

Satzbehalten in einer Zweitsprache: Anteile automatisierter und kontrollierter Verarbeitung

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Philosophie (Dr. phil.)
der

Erziehungswissenschaftlichen Fakultät
der Universität Erfurt

vorgelegt von
Dipl.-Psych. Sandra Kuhn

Erfurt 2011

Erstes Gutachten: Prof. Dr. Ralf Rummer (Universität Erfurt)

Zweites Gutachten: Prof. Dr. Gerd Mannhaupt (Universität Erfurt)

Drittes Gutachten: Prof. Dr. Thomas Lachmann (Techn. Universität Kaiserslautern)

Tag der Disputation: 03.02.2012

Tag der Promotion: 03.02.2012

urn:nbn:de:gbv:547-201200015

Zusammenfassung

Die Aufgabe "Sätze nachsprechen" wird in verschiedenen Sprachtests für Kinder im Vorschulalter verwendet, teilweise (z.B. bei der Sprachstandserhebung Delfin 4 in NRW) auch mit Kindern mit Deutsch als Zweitsprache.

Jedoch gibt es Hinweise darauf, dass die Sprachverarbeitung von Sprechern einer Zweitsprache weniger automatisiert verläuft als in einer Muttersprache, auch wenn die Zweitsprachler diese auf muttersprachlichem Niveau (u.a. fließend und akzentfrei) beherrschen und bereits in frühem Alter erworben haben.

Diese Unterschiede in der Verarbeitung wirken sich auf die Leistung in der Satz wiedergabe aus. In drei Studien mit Kindern im Vorschulalter, Schülern und jungen Erwachsenen mit Deutsch als Muttersprache bzw. Zweitsprache kann in dieser Arbeit ein Zusammenhang zwischen dem Sprachstatus (Muttersprache, Zweitsprache) und der Leistung in der Satz wiedergabe aufgezeigt werden. Dieser Leistungsunterschied kann in dieser Arbeit nicht in gleicher Form in anderen Maßen der Sprachkompetenz gefunden werden.

Der Unterschied wird daher als ein systematischer Nachteil für Zweitsprachler in der Bewertung ihrer Sprachkompetenz durch die Aufgabe "Sätze nachsprechen" gedeutet.

Die Betonung der Automatisierungsunterschiede wird auf die hohe Belastung des Arbeitsgedächtnisses während der Aufgabe zurückgeführt. Daher wird im zweiten Teil der Arbeit getestet, ob die inhaltliche (statt der wörtlichen) Wiedergabe der Sätze eine geringere Belastung mit sich bringt. Dies kann in Studie 4 für Muttersprachler gezeigt werden.

Jedoch führt auch die Beschränkung der Anforderung auf das inhaltliche Wiedergeben der Sätze nicht zu einer Angleichung der Wiedergabeleistung der Muttersprachler und Zweitsprachler mit ansonsten vergleichbarer Sprachkompetenz.

Es wird auf Grundlage der Ergebnisse dieser Arbeit davon abgeraten, die Aufgabe "Sätze wiederholen" als Indikator der Sprachkompetenz für Sprecher einer Zweitsprache im Vergleich mit Muttersprachlern zu verwenden.

Mehrsprachigkeit, Bilingualität

Verbales Arbeitsgedächtnis

Sprachtests, Sprachstandserhebungen

Satzbehalten, Sätze nachsprechen

Delfin 4

Abstract

Verbatim sentence repetition is a paradigm widely used to test the language competence of adult and child L1 and L2 speakers. The task is seemingly well suited for the purpose since it involves both comprehension and production of sentences. It is argued that a good knowledge of syntax is required to perform well, since the sentences contain more words than can be held in short term memory. However, from prior work on working memory and sentence recall it is known that attentional demands and the automaticity of sentence processing strongly influence performance in this task. Also, there is empirical evidence indicating that sentence processing may be less automatized (at least at some levels) even in highly proficient L2 speakers (near-natives) than it is in L1 speakers.

The empirical evidence is supported by current theories which model sentence recall as a combination of sentence comprehension and production, relying on limited attentional resources. More attentional resources are needed for sentence recall if sentence processing is automatized to a lesser degree. This results in a lower performance at this task due to the lower automaticity of sentence processing, even if language competence is native-like. Therefore, this task seems not suited to compare native's and near-native's language competence. The ability of near-natives may be systematically underestimated by this task.

It is argued that what is measured with sentence repetition tests is not determined by language ability alone, but determined by the degree of automaticity of sentence processing.

Studies 1-3 tested adult and child groups of native and near-native speakers of German with the tasks sentence repetition (auditory and visual presentation) and alternative tests of language competence, assessing language ability. Near-natives performed significantly lower in sentence recall than natives.

In studies 4 and 5, an alternative task (gist recall) is tested for its suitability for testing in a second language. Though gist recall seems to involve controlled processing to a lesser degree than verbatim recall (study 4), this test is also not well suited to compare the performance of speakers with an L2 to that of L1 speakers.

These findings have strong implications for models of L2-sentence processing and verbal working memory, as well as for language testing.

Bilingualism, Near-Native

Language testing, verbal working memory

Sentence repetition, sentence recall

Danksagung

Viele Menschen haben mich bei meinem Promotionsvorhaben fachlich und moralisch, direkt und indirekt unterstützt. Ich möchte die Gelegenheit nutzen, einigen von ihnen hier zu danken.

An erster Stelle danke ich Ralf Rummer, der mich während meiner Promotionszeit betreut hat und ohne den diese Arbeit nicht möglich gewesen wäre. Seine Leidenschaft für Psychologie und Sprache, seine Begeisterung für Forschung und die Förderung von Nachwuchswissenschaftlern, seien sie Studenten oder Promovenden, haben mich bereits während meines Studiums inspiriert. Ich hätte mir keinen besseren Mentor für mein wissenschaftliches "Erwachsenwerden" wünschen können.

Judith Schweppe danke ich für fachliche Diskussionen, für ihre Anteilnahme, ihre Klugheit und Hilfsbereitschaft und dafür, auch im chaotischen Unialltag einfach immer ansprechbar zu sein.

Gerd Mannhaupt danke ich für seine Unterstützung bei der Suche nach teilnehmenden Kindergärten, für seine Bereitschaft, diese Arbeit zu begutachten und ihm und seiner Familie für ihre Gastfreundschaft.

Thomas Lachmann danke ich ebenfalls im Voraus für die Bereitschaft, diese Arbeit zu begutachten.

Während der Arbeit an meiner Promotion wurde ich finanziell und ideell von der Stiftung der Deutschen Wirtschaft (sdw) unterstützt, herzlichen Dank hierfür.

Die hier berichteten Experimente sind Vorstudien zu bzw. Teil eines 2011 von der DFG bewilligten Projekts (Projektleiter sind Ralf Rummer und Judith Schweppe, Ru891/6-1).

Auch beim Forschungsreferat der Uni Erfurt möchte ich mich für die gewährte Unterstützung herzlich bedanken.

Mein herzlicher Dank gilt auch Anne Fürstenberg, die Teile dieses Textes gelesen und mir mit hilfreichen Anmerkungen Denkanstöße gegeben hat, und Antje Zindler, die mir bei der Programmierung von Studie 4 geholfen hat, sowie der gesamten Forschergruppe "Sprachbeherrschung", die mich mit wertvollen Kommentaren und fachlicher Expertise unterstützt hat.

Ich danke außerdem allen Kindergärten und Schulen, die bereit waren, meine Arbeit mit ihrer Teilnahme an den Studien zu unterstützen.

Für ihre Unterstützung in den nicht-fachlichen Aspekten des Doktoranden-Daseins, das (immer wieder) herzliche Willkommen in Erfurt und die ausgezeichnete Zusammenarbeit in verschiedenen Projekten, die meine Zeit in Erfurt bereichern haben, danke ich Mareike, Kerstin, Anne, Jessi, Judith, Julia, Julia, Christian, Angelika, Thomas und Wencke.

Meinen Eltern danke ich für ihre Unterstützung während meines Studiums und für ihr Vertrauen und Interesse während meiner Promotionszeit.

Lieber Daniel, Dir danke ich für Deine emotionale und fachliche Unterstützung, dafür dass Du geduldig alle Inhalte dieser Arbeit nicht nur gelesen, sondern auch wieder und wieder angehört (mitsamt meinen Sorgen und Nöten) und mit mir diskutiert hast und für Deine Hilfe in grafischen Dingen. Am meisten aber danke ich Dir dafür, dass Du mir die Kraft gibst, die Dinge umzusetzen, die ich mir vornehme und mich in meinen zahlreichen Projekten so bedingungslos unterstützt.

Inhaltsverzeichnis

Danksagung.....	I
Vorbemerkungen.....	1
1 Einleitung.....	9
1.1 Allgemeine Fragestellung.....	9
1.2 Gliederung.....	11
2 Sprachkompetenz und Automatisierung.....	14
2.1 Modelle von Sprachkompetenz.....	14
2.2 Automatisierung und Kontrolle.....	17
2.3 Grade der Automatisierung von Sprachverarbeitung.....	18
3 Mehrsprachigkeit.....	20
3.1 Erwerbsalter und maximal erreichbare Sprachfertigkeit.....	22
3.2 Automatisierung.....	24
4 Satzbehalten.....	32
4.1 Anwendungsfelder der Satzwiedergabe.....	32
4.2 Einfluss zentraler exekutiver Prozesse auf das Satzbehalten.....	35
4.3 Satzbehalten.....	41
5 Satzbehalten in einer Zweitsprache - Forschungsfrage und Hypothesen.....	52
6 Studie 1 - Sätze nachsprechen bei Vierjährigen.....	57
6.1 Methode.....	58
6.2 Ergebnisse.....	62
6.3 Diskussion.....	66
7 Studie 2 - Satzbehalten bei Erwachsenen mit L1 / L2 Deutsch.....	69
7.1 Methode.....	70
7.2 Ergebnisse.....	76
7.3 Diskussion.....	77
7.4 Gemeinsame Diskussion der Studien 1 und 2.....	79
8 Studie 3 - Satzbehalten bilingualer Schüler in Deutsch und Französisch.....	82
8.1 Methode.....	82
8.2 Ergebnisse.....	85
8.3 Diskussion.....	88
8.4 Gemeinsame Diskussion Studien 1-3.....	89
9 Studie 4 - Wörtliches und inhaltliches Satzbehalten in der L1.....	92
9.1 Methode.....	95
9.2 Ergebnisse.....	96
9.3 Diskussion.....	99
10 Studie 5 - Wörtliches und inhaltliches Satzbehalten in L1 / L2 Deutsch.....	101
10.1 Methode.....	101
10.2 Ergebnisse.....	104
10.3 Diskussion.....	109

10.4	Gemeinsame Diskussion der Studien 4 und 5	111
11	Allgemeine Diskussion.....	115
11.1	Annahmen und Ergebnisse im ersten Teil	115
11.2	Annahmen und Ergebnisse im zweiten Teil	115
11.3	Einschränkungen und Probleme der Studien 1, 2 und 3	117
11.4	Allgemeine Diskussion der Bedeutung der Ergebnisse für Theorie und Praxis.....	120
11.5	Weiterführende Forschungsfragen.....	124
12	Schlusswort.....	128
	Literatur.....	129
	Anhang.....	139
	Erklärung.....	165

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung des Mehrkomponentenmodells des Arbeitsgedächtnisses von Baddeley (2000).....	36
Abbildung 2: Schematische Darstellung des Modells des Arbeitsgedächtnisses als aktiviertes Langzeitgedächtnis (vgl. Cowan, 1988, 1993, 1999). Die einzelnen Punkte stellen Repräsentationen im Gedächtnis dar, die sich in unterschiedlichen Aktivierungszuständen befinden: nicht aktiviert, aktiviert und nicht im Fokus der Aufmerksamkeit, aktiviert und im Fokus der Aufmerksamkeit.	37
Abbildung 3: Schematische Darstellung der stufenweisen Verarbeitung von Sprache.	42
Abbildung 4: Schematische Darstellung des Rahmenmodells zum Satzbehalten von Rummer und Schweppe (2009).....	45
Abbildung 5: Vergleich der mittleren Leistung (als Prozent der maximal erreichbaren Punkte) von Vierjährigen mit DaM und DaZ für die Untertests 1, 3 und 4 des SETK 3-5 und den Untertest 2 "Satzgedächtnis".	63
Abbildung 6: Vergleich der mittleren Leistung (als Prozent der maximal erreichbaren Punkte) von Vierjährigen mit DaM und DaZ für die Untertests 1, 3 und 4 des SETK 3-5 und der Aufgabe "Sätze nachsprechen" aus Delfin 4.....	64
Abbildung 7: Vergleich der durchschnittlichen Leistung von Probanden mit Deutsch als L1 bzw. L2 im C-Test und in der Satzwiedergabe (als Prozent der korrekt wiedergegebenen Wörter pro Satz): Parallelisierung entsprechend C-Test-Leistung.....	76
Abbildung 8: Vergleich der durchschnittlichen Leistung von Probanden mit Deutsch als L1 bzw. L2 im C-Test und in der Satzwiedergabe (als Prozent der korrekt wiedergegebenen Wörter pro Satz): Parallelisierung entsprechend Leistung in der Satzwiedergabe.....	77
Abbildung 9: Vergleich der Leistung bilingualer Schüler mit L1 Deutsch bzw. L1 Französisch im deutschen C-Test und der Wiedergabe deutscher Sätze (als Prozent der korrekt wiedergegebenen Wörter).....	85
Abbildung 10: Vergleich der Leistung bilingualer Schüler mit Französisch als L1 bzw. L1 Deutsch im französischen C-Test und der Wiedergabe französischer Sätze (als Prozent der korrekt wiedergegebenen Wörter).....	86
Abbildung 11: Vergleich der Leistung bilingualer Schüler mit Französisch als L1 bzw. Deutsch als L1 in der Wiedergabe französischer und deutscher Sätze (als Prozent der korrekt wiedergegebenen Wörter).....	87
Abbildung 12: Vergleich der mittleren Lesezeit pro Wort (in ms) zwischen den drei Bedingungen wörtliches und inhaltliches Behalten und Verstehen von Sätzen.	97
Abbildung 13: Vergleich der mittleren Lesezeit pro Wort (in ms) zwischen den drei Bedingungen wörtliches und inhaltliches Behalten und Verstehen von Sätzen zum Satzanfang und Satzende.	98
Abbildung 14: Vergleich der erreichten Punktzahl im C-Test und der wörtlichen Satzwiedergabe für die Gruppen mit der Instruktion zur wörtlichen Satzwiedergabe.....	105

Abbildung 15: Vergleich der erreichten Punktzahl im C-Test und der inhaltlichen Satz wiedergabe (Kategorien "korrekt" und "Paraphrase" summiert) für die Gruppen mit der Instruktion zur inhaltlichen Satz wiedergabe. 105

Abbildung 17: Vergleich der Leistung in der wörtlichen Satz wiedergabe zwischen den Sprachgruppen (L1 Deutsch vs. L2 Deutsch) und Instruktionsbedingungen (wörtliche vs. inhaltliche Satz wiedergabe). 108

Tabellenverzeichnis

Tabelle 1: Die Abgrenzung automatischer und kontrollierter Informationsverarbeitungsprozesse (übernommen aus Rummer, 1996).....	17
Tabelle 2: Leistung der Kinder mit DaM und DaZ in den Untertests des SETK 3-5 und Delfin 4 "Sätze nachsprechen".....	63
Tabelle 3: Standardisierte Regressionskoeffizienten und Signifikanz für multiple Regression, abhängige Variable: Leistung in der Satzwiedergabe (SETK 3-5, Untertest 2, Teil 1).....	65
Tabelle 4: Übersicht über die Leistung der beiden muttersprachlichen Gruppen und der <i>Near-Natives</i> in den beiden Aufgabentypen.	74
Tabelle 5: Balance der Sprachen und Aufgaben im Design von Studie 3.	84
Tabelle 6: Leistungen in der deutschen Satzwiedergabe und dem deutschen C-Test für Schüler mit L1 Französisch und L1 Deutsch.....	85
Tabelle 7: Leistungen in der deutschen Satzwiedergabe und dem deutschen C-Test für Schüler mit L1 Französisch und L1 Deutsch.....	86
Tabelle 8: Mittlere Lesezeiten pro Wort in den drei Gruppen zu Satzbeginn und Satzende.	96
Tabelle 9: Mittlere Wiedergabeleistung (SD in Klammern) in Prozent und Leistung im C-Test für die Bedingungen und Sprachgruppen.	104

„Die Muttersprache in Deutschland ist Deutsch.“ (Koalitionsvertrag CDU / FDP in Nordrhein-Westfalen, Juni 2005, S. 53)

Vorbemerkungen

Sprachstandserhebungen in Deutschland

Die hier behandelten Forschungsfragen stehen im Kontext der derzeitigen bildungspolitischen Bemühungen um die Erfassung und Förderung der Sprachfähigkeit von Kindern vor der Einschulung. Sie wollen in diesem Zusammenhang einen Beitrag zur gesellschaftlichen Debatte um Sprachstandserhebungen leisten. Daher soll in diesen Vorbemerkungen auf die Diskussion zu den Sprachstandserhebungen in Deutschland, insbesondere auf das Verfahren Delfin 4 (Nordrhein-Westfalen) eingegangen werden. Um die Bedeutung herauszustellen, die dabei der Berücksichtigung der Situation von Kindern mit Deutsch als Zweitsprache (DaZ) zukommt, soll kurz die Migrationssituation in Deutschland umrissen werden. Im Anschluss daran wird auf die allgemeine Kritik an den derzeit eingesetzten Sprachstandsverfahren eingegangen. Schließlich werden grundsätzliche Kritikpunkte am Einsatz bestimmter Aufgabentypen mit Kindern mit DaZ aus Sicht der Gedächtnispsychologie und Bilingualismusforschung thematisiert.

Im Jahr 2007 hatten in Deutschland 8,8 % der Wohnbevölkerung¹ eine andere Staatsbürgerschaft als die deutsche. Dies greift jedoch in der Beschreibung der Migrationssituation zu kurz, weshalb im Mikrozensus des Statistischen Bundesamtes zusätzlich die Anzahl der Menschen mit Migrationshintergrund erhoben wird. Die Gruppe der Menschen mit Migrationshintergrund wird definiert als „alle nach 1949 auf das heutige Gebiet der Bundesrepublik Deutschland zugewanderten, sowie alle in Deutschland geborenen Ausländer und alle in Deutschland als Deutsche geborenen mit zumindest einem zugewanderten oder als Ausländer in Deutschland geborenen Elternteil.“ Dies waren im Jahr 2007 15,4 Millionen Menschen (ein Bevölkerungsanteil von 18,7 %) von diesen wurden rund ein Drittel in Deutschland geboren (vgl. Migrationsbericht 2008, Bundesamt für Migration und Flüchtlinge, 2010). Für einen großen Teil dieser Gruppe ist Deutsch nicht Muttersprache.

¹ Dieser Begriff meint die in Deutschland ansässige Bevölkerung. Er steht im Gegensatz zu einem Begriff von Bevölkerung, der auch deutsche Staatsbürger, die im Ausland leben, mit einbeziehen würde.

Der Koalitionsvertrag der derzeitigen Regierungsparteien CDU/CSU und FDP vom Oktober 2009 formuliert als ein bildungspolitisches Ziel, dass jedes Kind vor seiner Einschulung die deutsche Sprache beherrschen muss (vgl. Koalitionsvertrag CDU/CSU und FDP, 28.10.2009, S. 59). Als Mittel, dies zu erreichen, unterstützen die Regierungsparteien „verbindliche bundesweit vergleichbare Sprachstandstests für alle Kinder im Alter von vier Jahren.“ (Koalitionsvertrag CDU/CSU und FDP, 28.10.2009, S. 59-60). In allen Bundesländern mit Ausnahme Thüringens sind zum gegenwärtigen Zeitpunkt (Herbst 2011) Maßnahmen zu Erfassung des Sprachstandes von Kindergartenkindern getroffen worden. Dabei werden unterschiedliche Testverfahren und -prinzipien eingesetzt, so dass die Zielsetzung einer bundesweiten Vergleichbarkeit der Ergebnisse bisher nicht erreicht wurde. Teilweise wird der Sprachstand durch die ErzieherInnen eingeschätzt (z.B. Rheinland-Pfalz, Schleswig-Holstein), teilweise werden Sprachscreenings eingesetzt (z.B. Niedersachsen, Hessen). Einige Verfahren beinhalten ein ausgearbeitetes Sprachförderkonzept, werden wissenschaftlich begleitet und evaluiert und berücksichtigen die besondere Situation von Kindern mit DaZ, wie zum Beispiel das Programm „Sag' mal was“ (Baden-Württemberg) mit der "Linguistischen Sprachstandserhebung Deutsch als Zweitsprache", kurz LiSe-DaZ (Schulz & Tracy, in Druck). Das Hamburger Verfahren zur Analyse des Sprachstands bei Fünfjährigen (HAVAS 5, Reich & Roth, 2004) evaluiert neben dem Sprachstand in der deutschen Sprache auch die sprachliche Entwicklung in der Muttersprache der Kinder mit DaZ, während Delfin 4 (Diagnostik, Elternarbeit, Förderung der Sprachkompetenz in Nordrhein-Westfalen bei 4-Jährigen, Fried, 2007, 2008; seit 2010 auch in Sachsen-Anhalt im Einsatz) auf die Berücksichtigung der Erstsprache bei DaZ-Kindern verzichtet und auch in der Auswertung unberücksichtigt lässt, ob Deutsch Erstsprache (L1) oder Zweitsprache (L2) ist und wie viele Kontaktmonate die DaZ-Kinder zum Testzeitpunkt zu ihrer L2 Deutsch hatten. Einen umfassenden Überblick über die in Deutschland verfügbaren Sprachentwicklungstests und Sprachscreenings geben Kany und Schöler (2010).

Delfin 4

Die Sprachstandserhebungen, die derzeit in den einzelnen Bundesländern eingesetzt werden, werden aus verschiedenen Richtungen und mit unterschiedlichen Begründungen kritisiert. Neben ideologischer Grundsatzdebatten finden sich in der Tagespresse Diskussionen zu allgemeinen Problemen in der Durchführung der Tests und

die Frage nach der Finanzierung von Testverfahren und Sprachförderung, deren Sinnhaftigkeit und Effektivität (z.B. Artikel "Lernziel Deutsch", Felix Berth, Süddeutsche Zeitung, 23.06.2008; "Zu kurz, zu spät, zu abstrakt", Martin Spiewak, Die Zeit, 22.10.2010). Insbesondere die Einführung von Delfin 4 (Fried, 2007, Nordrhein-Westfalen) wurde mit großer Aufmerksamkeit und Kritik durch die direkt betroffene und fachlich interessierte Öffentlichkeit begleitet (vgl. Kany & Schöler, 2010). Die Internationale Vereinigung der Waldorfkindergärten sammelte binnen 5 Tagen 1552 Unterschriften von Eltern unter ihrem Aufruf „Jetzt reicht's – stoppt das Sprachstandsfeststellungs-Verfahren“; die regionalen Medien berichteten mehrfach über gravierende Probleme bei der Durchführung von Delfin 4 (z.B. „Bei Sprachförderung Ziel verfehlt“, Westfälische Rundschau vom 20.01.2010), und die Gewerkschaft für Erziehung und Wissenschaft (GEW) kritisierte das Verfahren in mehreren Pressemitteilungen (z.B. „Delfin Sprachtest geht baden“ 05.06.2008, „Tests sofort stoppen“ 26.04.2007, s. Anhang E2 und E3).

Fachliche Kritik an den Sprachstandserhebungen allgemein äußern die Deutsche Gesellschaft für Sprachheilpädagogik (dgs) und der Deutsche Bundesverband der akademischen Sprachtherapeuten (dbs) in einer gemeinsamen Stellungnahme: „Die derzeit in Deutschland Anwendung findenden Verfahren zur Feststellung des Sprachstandes z.B. DEUTSCH PLUS (Berlin), HAVAS 4 und 5 (Hamburg), FIT IN DEUTSCH (Niedersachsen), Delfin 4 (Nordrhein-Westfalen), SISMIK (Bayern) – genügen unseren Ansprüchen an ein aussagekräftiges Verfahren zur vorschulischen Sprachstandserhebung nicht.“ (Stellungnahme dgs/dbs, Oktober 2007, s. Anhang E 4). Auch aus den Universitäten erfahren die Erhebungen Kritik. Hasselhorn (2007) nennt in einer Expertise im Auftrag des Ministeriums für Kultus, Jugend und Sport Baden-Württemberg vom August 2007 sechs zentrale Klassen von Kompetenzen (in Anlehnung an Weinert, 2006), die von Sprachstandserhebungen getestet werden sollten, nämlich prosodisch-rhythmische Kompetenzen, phonologisch-phonetische Kompetenzen, morphologische Kompetenzen, syntaktische Kompetenzen, lexikalisch-semantische Kompetenzen und pragmatische Kompetenzen. Er stellt fest, dass nur wenige Sprachtests diese Kompetenzen valide erfassen, unter ihnen der SETK 3-5 (Sprachentwicklungstest für Kinder im Alter von 3 bis 5 Jahren, Grimm, 2001; vgl. Anhang E1: Hasselhorn, Expertise über derzeit verfügbare wissenschaftlich fundierte Verfahren einer umfassenden Sprachstandsdiagnostik bei Vier- und Fünfjährigen, 13.08.2007).

Diesen berechtigten Kritikpunkten soll in dieser Arbeit ein weiterer hinzugefügt werden. Er betrifft die Angemessenheit der Aufgabenstellung eines (von insgesamt vier) Untertests von Delfin 4 und dem SETK 3-5 für die Gruppe der Kinder mit DaZ. Im Gegensatz zum SETK 3-5, der für eine monolinguale Zielgruppe mit L1 Deutsch entwickelt wurde und auch nur für die Verwendung mit dieser empfohlen wird, ist Delfin 4 seit seiner Einführung auch für den Einsatz mit DaZ-Kindern vorgesehen. Dabei verwendet Delfin 4 teilweise ähnliche Aufgabentypen wie der SETK 3-5, so zum Beispiel den Aufgabentyp "Sätze nachsprechen" (Delfin 4, Stufe 1), der dem Aufgabentyp "Satzgedächtnis" (SETK 3-5) entspricht. Dies geschieht ohne weitere Überprüfung der Angemessenheit dieses Aufgabentyps für Kinder mit DaZ und ohne Normierung an einer Stichprobe von DaZ-Kindern. Kany und Schöler (2010) halten zwar einzelne SETK 3-5 Untertests für geeignet für den Einsatz mit Kindern mit Migrationshintergrund; der Untertest "Satzgedächtnis" gehört jedoch nicht dazu. Obwohl die Autoren dies an dieser Stelle nicht weiter begründen, erscheint ihre Einschätzung vor dem Hintergrund der aktuellen Forschung zum Satzbehalten und zur Mehrsprachigkeit stimmig. So gibt es theoretische und empirische Hinweise auf fundamentale Unterschiede im Satzbehalten in einer L1 und einer L2, bedingt durch unterschiedliche Grade der Automatisierung der Sprachverarbeitung in L1 und L2. Dieser Zusammenhang wird in den nächsten Kapiteln ausführlich dargestellt.

Der unkritische Umgang mit Maßen der Sprachkompetenz im Zusammenhang mit einer Zweitsprache hat den Anstoß zu dieser Arbeit gegeben. Die generelle Ausblendung der Forschung zur Mehrsprachigkeit und zum Satzbehalten, insbesondere in der Entwicklung des Sprachscreenings Delfin 4, Stufe 1, das auch mit Kindern mit DaZ verwendet wird, ist weder aus wissenschaftlicher noch aus gesellschaftlicher Perspektive nachvollziehbar. Um den Kontext der Einführung von Delfin 4 in NRW und das Verfahren selbst angemessen zu beschreiben, möchte ich diese im Folgenden detaillierter beschreiben und auf Probleme in Design und Durchführung eingehen, die außerhalb meiner eigenen Fragestellung liegen. Diesen schon bestehenden Kritikpunkten möchte ich dann meine fachliche Kritik im Zusammenhang mit der Aufgabe "Sätze nachsprechen" hinzufügen.

Delfin 4: Konzeption und Durchführung

Delfin 4, Stufe 1 „Besuch im Zoo“ ist als Brettspiel mit Spielfiguren in vier Farben, einem Würfel und Spielkarten mit Tiermotiven gestaltet. Vier Kinder nehmen gleichzeitig

teil, das Spiel wird von einer/m ErzieherIn angeleitet, eine Lehrkraft aus der betreuenden Grundschule protokolliert die Antworten der Kinder. Das Spielbrett symbolisiert einen Spaziergang durch einen Zoo, einzelne Tierkäfige sind Stationen für die Spielfiguren, an denen Aufgaben bearbeitet werden müssen. Insgesamt erhält jedes Kind vier Aufgaben aus folgenden Kategorien: a) Nachsprechen von Sätzen, b) Nachsprechen von Nichtwörtern, c) Ausführen von Handlungsanweisungen, d) Bildbeschreibung. Nach der ersten Stufe wird anhand der erreichten Gesamtpunktzahl entschieden, ob die Kinder Sprachförderung erhalten werden oder nicht, beziehungsweise ob sie zur Klärung dieser Frage an der zweiten Stufe des Verfahrens, dem „Besuch im Pfiffikushaus“, teilnehmen müssen. Die Verpflichtung zur Teilnahme an der Sprachstandserhebung wurde im Schulgesetz verankert.

Die Durchführung in 4er-Gruppen soll einen ökonomischen Ablauf ermöglichen, bringt aber entscheidende Nachteile im Vergleich zu Einzelsitzungen mit sich. Die Kinder nehmen während zwei Dritteln bzw. drei Vierteln der Testzeit (je nachdem, ob drei oder vier Kinder gleichzeitig getestet werden, beides ist vorgesehen), die etwa 45 Minuten in Anspruch nimmt, nicht aktiv teil. Daher dauert der Test für das einzelne Kind unnötig lange und ist damit ermüdend; Konzentration und Teilnahmebereitschaft sinken. Die vorgegebene Sitzordnung beim Test führt dazu, dass die Protokollantin mindestens ein Kind nur von hinten sieht, was die Beurteilung erschwert, da visuelle Informationen zur Artikulation nicht zur Verfügung stehen. Zwischenrufe und das „Vorsagen“ von Kindern, die nicht an der Reihe sind, können die Ergebnisse verfälschen, da keine Ersatzfragen für einen solchen Fall vorgesehen sind. Daneben ist es zweifelhaft, ob alle Aufgabentypen für die getesteten Altersstufen angemessen sind. Da Delfin 4, Stufe 1 im Frühjahr durchgeführt wird, also nahezu zweieinhalb Jahre vor der Einschulung, hat ein Teil der getesteten Kinder das vierte Lebensjahr noch nicht vollendet. Der Altersunterschied wird in der Auswertung zwar durch unterschiedliche Normen für Kinder unter bzw. über 4 Jahren berücksichtigt. Diese Maßnahme ist jedoch vielleicht nicht ausreichend, da es fraglich ist, ob der Untertest "Nachsprechen von Sätzen" für die Altersgruppe der unter 4-Jährigen überhaupt geeignet sind. Ein Indiz für letztere Annahme ist, dass andere Sprachentwicklungstests für Kinder auf den Einsatz dieses Aufgabentyps für unter 4-Jährige verzichten, so zum Beispiel der SETK 3-5 (Grimm, 2001; vgl. Kany & Schöler, 2010). Letzterer ist ein seit vielen Jahren etabliertes Testwerkzeug, in dessen Entwicklung außerdem die jahrelange Erfahrung der Erstautorin in Entwicklung und Überprüfung des HSET (Heidelberger Sprachentwicklungstest, Grimm & Schöler, 1978, 2. Auflage 1991)

einfließen. Er verwendet eine unterschiedliche Auswahl von Untertests für Kinder unter beziehungsweise über vier Jahren. Der Untertest "Satzgedächtnis", der der Aufgabe "Nachsprechen von Sätzen" in Delfin 4 entspricht, wird beim SETK 3-5 erst für Kinder ab 4 Jahren eingesetzt.

Der durch Delfin 4, Stufe 1 festgestellte Sprachförderbedarf ist regelmäßig sehr hoch: 2008 wurden von rund 160 000 Kindern in der ersten Stufe des zweistufigen Testverfahrens rund 37 000 als förderbedürftig klassifiziert, also 23 % (vgl. Schulministerium NRW, Feststellung des Sprachstands zwei Jahre vor der Einschulung, Fachinformation zum Verfahren 2009). Es ist fraglich, ob die Anforderungen an die Sprachfähigkeiten, die Delfin 4 an Kinder stellt, die zweieinhalb Jahre vor der Einschulung stehen, angemessen sind, wenn 23 % der Kinder diesen Anforderungen nicht genügen. Dieses Ergebnis legt eine Überprüfung des angelegten Maßstabs und der Messgenauigkeit des Tests nahe². Daneben ist zu bemängeln, dass kein Förderkonzept entwickelt wurde, das auf den Ergebnissen der Diagnose aufbaut. Die Förderung, die sich an den Test anschließt, bleibt in der Verantwortung des einzelnen Kindergartens bzw. der einzelnen Kindertagesstätte, die lediglich einen finanziellen Zuschuss von 340 € pro Jahr und Kind mit Sprachförderbedarf erhalten. Bisher gibt es noch keine Evaluation des Einflusses von Delfin 4 (und der darauf aufbauenden finanziellen Unterstützung der Kindergärten und -tagesstätten) auf die Sprachkenntnisse zum Zeitpunkt der Einschulung. Somit kann bislang nicht beantwortet werden, ob das Verfahren das gesetzte Ziel erreicht, die Sprachfertigkeiten der Kinder zum Zeitpunkt der Einschulung zu verbessern. Eine Evaluation ist für den Einschulungsjahrgang 2011/12 geplant.

Ein weiteres gravierendes Problem stellt die eingeschränkte Übertragbarkeit der Normen von Delfin 4 auf Kinder mit DaZ dar. Die Normierung von Delfin 4 erfolgte an einer Teilstichprobe des ersten Durchgangs (Erhebung 2007) von 14 859 Kindern (Fried, Briedigkeit, Isele & Schunder, 2009). Damit wurde der Test an einer außergewöhnlich großen Teilstichprobe normiert, die allerdings ausschließlich aus Kindern mit Deutsch als Muttersprache (DaM) bestand. Die Bewertung von Kindern mit DaZ auf Grundlage der Leistung von Kindern mit DaM, ohne die Berücksichtigung von Kontaktmonaten mit der L2, ist jedoch höchst problematisch. Begründet wird sie in der Handreichung zu Delfin 4

² Siehe hierzu auch die Unterscheidung von Kany und Schöler (2010) zu statistischen und sachlichen Bezugsnormen. Unter der Verwendung sachlicher Bezugsnormen wären diese Ergebnisse durchaus nicht erstaunlich. Jedoch berichten Fried et al. die Normierung an einer repräsentativen Stichprobe mit Auslassung der Daten nicht muttersprachlicher Kinder (vgl. Fried et al, 2009).

damit, dass „vorausgesetzt wird, dass auch Kinder mit nichtdeutscher Muttersprache hinreichende Deutschkenntnisse benötigen, damit sie zum Schulbeginn die gleichen Bildungschancen haben wie Kinder mit deutscher Muttersprache.“ (Fried, Briedigkeit & Schunder, 2009, S.1). Welcher Kenntnisstand eines vierjährigen Kindes mit DaZ ist jedoch Voraussetzung für seine sprachliche Schulreife im Alter von sechs Jahren? Um eine valide Prognose über den Verlauf der weiteren sprachlichen Entwicklung eines Kindes mit DaZ zu machen, darf die Leistung in einem für Kinder mit DaM normierten Verfahren nicht isoliert betrachtet werden. Die Aussagekraft, die das Verfahren für Vierjährige mit DaM bezüglich ihrer zu erwartenden Sprachkompetenz im Alter von 6 Jahren hat (die ebenfalls für Delfin 4 noch nicht nachgewiesen wurde), ist nicht gleichzusetzen mit der Aussagekraft, die es über die Entwicklung eines Kindes mit DaZ hat. Die Begründung des Verfahrens durch den erwünschten Zielzustand (gleiche Fertigkeiten und Chancen bei Schulbeginn) blendet die unterschiedlichen Entwicklungsvoraussetzungen zum Testzeitpunkt aus (allen voran die Kontaktmonate mit der Sprache) und verfälscht so die prognostische Aussagekraft der aktuellen Leistung bezüglich des Zustandes zur Einschulung. Letztlich muss es Absicht der Sprachstandserhebungen sein, kritische Verzögerungen im Spracherwerb aufzudecken, die einer gezielten Förderung bedürfen. Dazu muss sie diese Verzögerungen von der normalen Entwicklung in der Zweitsprache unterscheiden können, wozu es immer nötig ist, den Startzeitpunkt dieser Entwicklung, also den Start des regelmäßigen Kontakts zur L2, in die Beurteilung mit einzubeziehen. Es ist gut belegt, dass Kinder beim frühen Zweitspracherwerb (Erwerbstart zwischen 1;6 und 4;0 Jahren) die gleichen Meilensteine durchlaufen wie beim Erstspracherwerb (vgl. De Houwer, 2009; Schulz, Tracy & Wenzel, 2008; Weinert, 2006). Jedoch muss ein Urteil über das Erreichen dieser Meilensteine neben dem Alter der Kinder immer auch die Anzahl der Kontaktmonate mit der Zweitsprache berücksichtigen (vgl. Schulz et al., 2008). Nur so kann bewertet werden, ob die Entwicklung verzögert verläuft und Förderung nötig und möglich ist. Schulz, Tracy und Wenzel (2008) konnten zeigen, dass in ihrer Stichprobe mehr als 80 % der Kinder mit DaZ nach 36 Kontaktmonaten alle Meilensteine erreicht hatten, die ein vierjähriges Kind mit DaM und normaler Sprachentwicklung erreicht hat. Nach 46 Kontaktmonaten hatten nahezu alle Kinder diese Meilensteine erreicht. Eine Fokussierung auf das Alter ohne Berücksichtigung der Kontaktmonate ist in der Bewertung der Sprachentwicklung von Kindern mit DaZ kontraproduktiv.

Zusätzlich zu der eingeschränkten Aussagekraft aufgrund der generellen Gleichbehandlung von Kindern mit DaZ und DaM bei der Auswertung der Ergebnisse besteht der Verdacht auf eine weitere Benachteiligung der DaZ-Kinder durch die Auswahl der Aufgabentypen in Delfin 4. Die Wiedergabe von Sätzen, die in der ersten Stufe des Verfahrens als einer von vier Aufgabentypen ("Sätze nachsprechen") verwendet wird, ist vermutlich nicht geeignet, die Sprachfähigkeit kompetenter Kinder (und Erwachsener) mit L2 Deutsch abzubilden. Aktuelle Modelle und empirische Forschungsergebnisse, sowohl zum Satzbehalten als auch zur Verarbeitung einer L2, lassen insbesondere Zweifel daran aufkommen, ob die Satz wiedergabe ein geeignetes Maß ist, um die Sprachkompetenz von Muttersprachlern und Sprechern einer L2 zu vergleichen³. Zahlreiche Studien zur Mehrsprachigkeit deuten auf eine generell geringere Automatisierung einer L2 im Vergleich zu einer L1 hin (z.B. Birdsong, 2006; Stowe & Sabourin, 2005). Dieser unterschiedliche Automatisierungsgrad ist ab einem bestimmten (sehr hohen) Grad der erreichten Sprachkompetenz in der L2 für den alltäglichen und akademischen Sprachgebrauch ohne Bedeutung. Er könnte jedoch bedeutsam werden unter der ungewöhnlich hohen Belastung des Arbeitsgedächtnisses, die bei der Aufgabe "Sätze wiedergeben" vorliegt (vgl. Aaronson & Scarborough, 1976). Dies könnte dazu führen, dass kompetente Zweitsprachler im Vergleich zu Muttersprachlern in dieser Aufgabe schlechter abschneiden, auch wenn sie über allgemein vergleichbare Sprachkompetenz verfügen. Für die Verwendung der Aufgabe "Sätze nachsprechen" mit Kindern mit DaZ würde dies bedeuten, dass die Sprachkompetenz nicht adäquat erfasst wird. Auch wenn zu erwarten ist, dass Kinder mit DaZ aufgrund ihres kürzeren Kontakts mit der deutschen Sprache durchschnittlich schlechter abschneiden als Kinder mit DaM, so wird dieser Unterschied durch die Testaufgabe "Sätze nachsprechen" zusätzlich künstlich überbetont. Dies betrifft vor allem Kinder mit DaZ, die ihre L2 Deutsch bereits sehr gut beherrschen. Ihre Sprachkompetenz wird durch das Testen mit der Satz wiedergabe unterschätzt. In dieser Arbeit untersuche ich, ob sich diese systematische Unterschätzung der Sprachkompetenz empirisch feststellen lässt und ob eine leichte Veränderung der Aufgabenstellung Abhilfe schaffen kann. Sollten sich die Thesen bestätigen, so kann nur empfohlen werden, die Aufgabe "Sätze nachsprechen" in ihrer derzeitigen Form nicht länger beim Testen von Kindern mit DaZ einzusetzen.

³ Diese Sprecher werden im Folgenden in Analogie zum Begriff Muttersprachler auch als Zweitsprachler bezeichnet.

1 Einleitung

1.1 Allgemeine Fragestellung

Die vorliegende Arbeit bewegt sich innerhalb zweier Spannungsfelder. Zum einen steht sie als interdisziplinäre Forschungsarbeit zwischen den Traditionen der Psychologie und denen der Linguistik. Zum anderen schlägt sie einen Bogen von der Grundlagenforschung bezüglich verbalem Arbeitsgedächtnis und Mehrsprachigkeit zur Anwendungsfrage der angemessenen Methoden zur Erfassung der Sprachkompetenz in einer Erst- und Zweitsprache. Dabei steht ein in der Sprachdiagnostik häufig verwendeter Aufgabentyp, das wörtliche Wiederholen von Sätzen im Mittelpunkt. Die Satz wiedergabe ist in einer Muttersprache ein guter Indikator für den Sprachentwicklungstand bei Kindern (vgl. Grimm, 2001; Kany & Schöler, 2010). Es ist jedoch nicht sicher, ob sie auch für Kinder, die Deutsch als Zweitsprache sprechen, eine valide Einschätzung ihrer Sprachkompetenz liefern kann. Zweifel daran kommen auf, wenn die Aussagen theoretischer Modelle und empirischer Hinweise sowohl zur Mehrsprachigkeit, als auch zur Satzverarbeitung und zum Arbeitsgedächtnis betrachtet werden. Zusammengefasst lässt sich aus diesen ableiten, dass das Satzbehalten in einer Zweitsprache (L2) sich von dem in einer Erstsprache (L1) unterscheidet. Die Satz wiedergabe in einer L2 betont vermutlich Verarbeitungsunterschiede, die nur unter außergewöhnlich hoher Belastung des Arbeitsgedächtnisses Einfluss auf die Leistung in einer sehr gut beherrschten L2 haben. Das Behalten eines Satzes stellt eine solch hohe Belastung dar, da bereits während des Lesens oder Hörens des Satzes und später während der Wiedergabe neben den Sprachverarbeitungsprozessen auch der Prozess des Behaltens bewältigt werden muss. Beide Prozesse greifen laut gängiger Modellannahmen (vgl. Just & Carpenter, 1992) auf eine geteilte Ressource zu. Wird eine größere Kapazität dieser Ressource während der parallel ablaufenden Prozesse der Verarbeitung und des Behaltens für die Verarbeitung benötigt, so bleibt weniger Kapazität für das Behalten, die Behaltensleistung sinkt. Eigentlich hochkompetente Sprecher werden in ihrer Sprachkompetenz durch diesen Aufgabentyp unterschätzt. Dies möchte ich in dieser Arbeit theoretisch darlegen und empirisch überprüfen.

Um dies zu leisten, werden in den ersten Kapiteln Konzepte und Evidenz aus den verschiedenen Forschungsfeldern vorgestellt. Es wird durchgängig versucht, die Verbindungen der einzelnen Teilgebiete zur Beantwortung der Fragestellung zu

verdeutlichen. Ich möchte der Zusammenführung dieser Teilgebiete hier kurz vorgreifen, auch wenn die Konzepte dahinter noch nicht erläutert wurden.

Die Verarbeitung von Sprache ist ein hochautomatisierter Prozess. Neben der Mühelosigkeit des Verstehens und Produzierens von Sprache spricht dafür auch die Unaufhaltsamkeit des Verstehensprozesses: Man kann eine beherrschte Sprache nicht hören, ohne sie "automatisch" zu verstehen. Das Behalten von sprachlichen Inhalten und insbesondere der Satzoberfläche zur Wiedergabe erfolgt jedoch nicht automatisch, sondern muss kontrolliert gesteuert werden. Die Kapazität für diese kontrollierten Prozesse ist begrenzt. Bei einer großen Menge von Inhalts- und Satzoberflächeninformation, die behalten werden muss, wird diese Kapazität nahezu ausgelastet⁴; im Falle einer vollständigen Automatisierung der Sprachverarbeitung wird sie gänzlich auf die Aufrechterhaltung von Satzinhalt und Satzoberfläche (z.B. Syntax, Morphologie, Wortwahl) verwendet.

In verschiedenen Studien konnte gezeigt werden, dass die Verarbeitung einer sehr gut beherrschten Zweitsprache weniger automatisiert verläuft, als die Verarbeitung einer Erstsprache, auch ohne dass dies sich zwangsläufig in der Kompetenz niederschlägt, mit der diese Zweitsprache beherrscht wird. Die Verarbeitung ist also in gewissem Sinne kognitiv aufwändiger, was aber weder subjektiv unmittelbar erfahrbar, noch objektiv in Kompetenzunterschieden messbar ist. Dieser Unterschied im Automatisierungsgrad zeigt sich in Abwesenheit von Leistungsunterschieden in bildgebenden Verfahren. Jedoch hat er für die Leistung in der Satzwiedergabe Konsequenzen. Da bei der Verarbeitung einer Zweitsprache mehr Ressourcen für die Verarbeitung an sich aufgewandt werden müssen, stehen diese nicht für die Behaltensaufgabe zur Verfügung. Die Leistung im Satzbehalten in einer L2 wird also bei entsprechender Länge der zu behaltenden Sätze unter der Leistung in der L1 bzw. unter der Leistung von Muttersprachlern liegen. Dennoch kann aus dieser niedrigeren Leistung (und aus dem sie wahrscheinlich bedingenden niedrigeren Automatisierungsgrad) nicht unbedingt auf eine niedrigere Sprachkompetenz⁵ geschlossen werden. Unterschiede zwischen Muttersprachlern und hochkompetenten Sprechern einer L2 (*Near-Natives*) im Automatisierungsgrad der Sprache sind zwar vorhanden. Insgesamt sind die *Near-Natives* jedoch auf einem so hohen Niveau der Sprachkompetenz und die Verarbeitung ihrer L2 ist zu einem so hohen Maße automatisiert, dass die verbleibenden

⁴ Im Falle von sehr langen Sätzen wird sie gar überlastet, was zu einem Verlust an Information führt.

⁵ Welche Definition von Sprachkompetenz dieser Arbeit zu Grunde liegt, wird ausführlich im nächsten Kapitel diskutiert.

Unterschiede zu Muttersprachlern irrelevant sind. Sie interessieren in experimentellen und allgemein in Forschungskontexten, sollten jedoch in einer für Sprachtests relevanten Definition von Sprachkompetenz nicht als Unterschiede gedeutet werden. Wie in den Vorbemerkungen bereits dargelegt, muss insbesondere ihre Aussagekraft in Sprachstandserhebungen mit Kindern mit DaZ im Vorschulalter angezweifelt werden.

1.2 Gliederung

Die theoretischen und empirischen Anhaltspunkte, die gegen den Einsatz der Aufgabe "Sätze nachsprechen" beim Testen in einer L2 sprechen, werden in den folgenden Kapiteln ausführlich behandelt. Zunächst folgt in Kapitel 2 eine Erläuterung der Konzepte Sprachkompetenz und Automatisierung, wie sie in dieser Arbeit verwendet werden. In Kapitel 3 folgt eine Übersicht über die Forschung zur Verarbeitung einer L2 mit dem Schwerpunkt auf der Automatisierung der Satzverarbeitung hochkompetenter Sprecher. Kapitel 4 diskutiert Aussagen gängiger Konzeptionen des verbalen Arbeitsgedächtnisses zum Satzbehalten. Kapitel 5 leitet aus diesen Ausführungen die Fragestellungen und Annahmen der Arbeit zum Satzbehalten in einer L2 ab. Darauf aufbauend werden in den folgenden Kapiteln fünf Studien vorgestellt, die diese Annahmen prüfen. Diese Studien teilen sich in einen ersten Teil (Studien 1-3) auf, in dem die grundlegenden Annahmen der Arbeit geprüft werden, und einen zweiten Teil (Studien 4 und 5), in dem eine alternative Aufgabenstellung auf ihre Tauglichkeit für das Testen von Sprachkompetenz in einer L2 untersucht wird.

Im ersten Teil (Studien 1-3) wird zunächst die Annahme überprüft, dass unter vierjährigen Kindern tatsächlich ein aufgabenspezifisches schlechteres Abschneiden im Nachsprechen von Sätzen beobachtet werden kann. Dazu wird die Leistung von Kindern mit DaZ im Nachsprechen von Sätzen verglichen mit ihrer Leistung in anderen Aufgabentypen eines gängigen Sprachentwicklungstests (SETK 3-5, Grimm, 2001) und der Leistung gleichaltriger Kinder mit DaM in den jeweiligen Aufgaben. Obwohl gängige Sprachtests für Kinder zwei Arten von Sätzen verwenden, sinnvolle und sinnlose, beschränkt sich meine Betrachtung auf die sinnvollen Sätze. Die Betrachtung der Verarbeitung sinnloser Sätze in einer Zweitsprache ist zu komplex, um sie im Rahmen dieser Arbeit zu behandeln. Die Verarbeitung semantischer Anomalien und die Unterschiede in der Verarbeitung dieser in einer L1 und L2 würden die Komplexität der

Forschungsfrage so weit erhöhen, dass sie im Rahmen dieser Arbeit nicht zu bewältigen gewesen wäre.

Die Kinder mit DaZ in Studie 1 haben jedoch den Erwerb ihrer L2 (und auch ihrer L1) noch nicht abgeschlossen. Wenn tatsächlich grundlegende Verarbeitungsunterschiede zwischen einer L1 und einer L2 für das aufgabenspezifisch schlechtere Abschneiden verantwortlich sind, so sollten diese auch nach Abschluss des erfolgreichen L2-Erwerbs noch vorhanden sein; sind sie das nicht, so sind die Unterschiede, die zwischen den Kindern aufgedeckt wurden, doch indikativ für den Stand ihrer Entwicklung in der L2. Um Letzteres auszuschließen wurden erwachsene hochkompetente Sprecher der L2 Deutsch getestet (Studie 2). Ihre Leistung in der Satzwiedergabe und in einem Vergleichsmaß der Sprachkompetenz, dem C-Test, wird wiederum mit der Leistung von erwachsenen Muttersprachlern in beiden Tests verglichen. Es wird untersucht, ob *Near-Natives*, die im Vergleichsmaß auf muttersprachlichem Niveau abschneiden, sich in ihrer Leistung in der Satzwiedergabe von Muttersprachlern unterscheiden. Daneben wird überprüft, ob dieses Ergebnis vielleicht auf die niedrigere Diskriminationsfähigkeit des C-Tests im Vergleich zur Satzwiedergabe zurückzuführen ist: Teilnehmer schneiden unterschiedlich in C-Test und Satzwiedergabe ab, weil die Satzwiedergabe in der Lage ist, feinere Unterschiede aufzudecken. Um dies auszuschließen, werden Muttersprachler mit den *Near-Natives* ähnlicher Leistung in der Satzwiedergabe in ihrer C-Test-Leistung mit den *Near-Natives* verglichen.

Der kritische Verarbeitungsunterschied ist ein Unterschied zwischen der Verarbeitung einer L1 und einer L2, nicht zwischen bilingualen Sprechern und monolingualen Sprechern. Daher soll in Studie 3, die den ersten Teil der Arbeit abschließt, gezeigt werden, dass der aufgabenspezifische Nachteil auch in einer insgesamt bilingualen Stichprobe jeweils für die jeweilige L2 zu finden ist. Da in Studie 3 in zwei Sprachen (Deutsch und Französisch) getestet wird, zeigt diese Studie auch die Übertragbarkeit der ersten Ergebnisse in eine andere Sprache auf.

Nachdem im ersten Teil Evidenz für die grundlegenden Annahmen gesammelt wurde, soll im zweiten Teil eine Alternative zur wörtlichen Satzwiedergabe untersucht werden. Da die Benachteiligung der L2-Sprecher vermutlich auf die hohe Belastung des Arbeitsgedächtnisses zurückzuführen ist, sollte sich diese Alternative in diesem Aspekt von der wörtlichen Satzwiedergabe unterscheiden. Aaronson und Scarborough (1976)

fanden in einer *self-paced*-Lesestudie Unterschiede zwischen der Arbeitsgedächtnisbelastung beim Lesen von Sätzen zum Verstehen und beim Lesen zum Behalten. Das Lesen zum Behalten ging mit einer höheren Belastung einher, wie Aaronson und Scarborough (1976) aus der geringeren Lesegeschwindigkeit beim Behalten im Vergleich zum (reinen) Verstehen schlossen. Wie weiter unten ausführlicher argumentiert wird, stellt insbesondere das Behalten der Satzoberfläche (genaue Syntax, Morphologie, Wortwahl) eine besondere und ungewöhnliche Belastung dar im Vergleich zum eher zu Kommunikationszwecken üblichen Behalten des Satzinhalts. Daher wird im zweiten Teil in Studie 4 zunächst überprüft, ob das Behalten des *Satzinhalts* (statt Inhalt *und* Oberfläche) eine geringere Belastung des Arbeitsgedächtnisses darstellt, als das wörtliche Behalten. Dazu wird dem Design der Lesestudie von Aaronson und Scarborough (1976) eine weitere Bedingung, nämlich die der inhaltlichen Wiedergabe, hinzugefügt und überprüft, ob die Belastung des Arbeitsgedächtnisses (operationalisiert durch die Messung der Lesegeschwindigkeit) beim inhaltlichen Behalten geringer ausfällt als beim wörtlichen Behalten.

In der letzten Studie (Kapitel 10) wird schließlich untersucht, ob der aufgabenspezifische Unterschied zwischen Sprechern einer L1 und L2 in der Satzwiedergabe im Vergleich zu einem anderen Maß der Sprachkompetenz verschwindet, wenn die Sätze inhaltlich statt wörtlich wiedergegeben werden. Trifft dies zu, so könnte die inhaltliche Satzwiedergabe als Alternative in Sprachtests die wörtliche Satzwiedergabe ersetzen. Abschließend werden in Kapitel 11 die Ergebnisse der Studien zu Modellen der bilingualen Sprachverarbeitung und des Satzbehaltens in Bezug gesetzt und mögliche weiterführende Forschungsfragen diskutiert.

2 Sprachkompetenz und Automatisierung

In diesem Kapitel werden die Konzepte Sprachkompetenz und Automatisierung und ihr Zusammenhang diskutiert. Dazu werde ich den hier verwendeten Sprachkompetenzbegriff zunächst von Chomskys Begriff der Kompetenz abgrenzen. Danach werden die Sprachkompetenzmodelle angerissen, die die Grundlage von Delfin 4 und dem SETK 3-5 bilden, und ihre Umsetzung in diesen Sprachtests diskutiert. Abschließend werde ich argumentieren, dass und warum Unterschiede in der Automatisierung nicht in jedem Fall als Unterschiede in der Sprachkompetenz gedeutet werden sollten.

2.1 Modelle von Sprachkompetenz

In der Sprachwissenschaft ist der Begriff Kompetenz eng mit den Theorien von Chomsky (z.B. 1965) verknüpft. Chomsky stellt dem Begriff der Kompetenz, als dem bewussten und unbewussten Wissen über eine Sprache, den der Performanz gegenüber, der das Sprachverhalten in den einzelnen Instanzen des Sprachgebrauchs (also z.B. die Äußerung eines bestimmten Satzes) beschreibt (vgl. Bußmann, 1990). Zwischen der Kompetenz (also dem Wissen über die Sprache) und der Performanz (also dem tatsächlichen sprachlichen Verhalten) stehen Prozesse der Sprachverarbeitung, die Chomsky nicht weiter beschreibt. Nach Chomskys Definition ist allein die Performanz beobachtbar und messbar, nicht aber die Kompetenz. Da in dieser Arbeit ein Konzept von Sprachkompetenz nötig ist, das prinzipiell messbar ist und außerdem die Sprachkompetenz im Zusammenhang mit der Verarbeitung von Sprache betrachtet werden soll, wird der Begriff der Sprachkompetenz hier losgelöst von Chomskys Begriff der Kompetenz verwendet.

Sprachstandserhebungen wollen den Sprachstand, also die sprachliche Kompetenz zu einem bestimmten Zeitpunkt, erfassen. Sprachtests für Kinder unterscheiden sich darin, welcher Begriff von Kompetenz ihnen zu Grunde liegt und darin, ob sie umfassend die allgemeine Sprachkompetenz oder ihre Teilkomponenten (z.B. Wortschatz, grammatisches Verständnis) erfassen wollen. Kany und Schöler (2010) grenzen in der Beschreibung der theoretischen Konzeption von Sprachstandstests allgemeine sprachliche Fähigkeiten von grundlegenden Funktionen der Informationsverarbeitung ab. Der SETK 3-5 will neben sprachlichen Fertigkeiten auch diese Funktionen der Informationsverarbeitung erfassen.

Dies hat die Erhöhung der prognostischen Funktion des Tests zum Ziel: "Bei den Sprach- und Schriftspracherwerbsstörungen gelten größtenteils Defizite in der auditiven Informationsverarbeitung als verursachend, insofern erhält die prognostische Funktion der Diagnostik durch den Einbezug von Gedächtnisaufgaben im SETK 3-5 ein größeres Gewicht" (Kany & Schöler, 2010, S. 151). Die phonologische Gedächtnisfunktion wird im SETK 3-5 in den Untertests "Satzgedächtnis", "Phonologisches Gedächtnis für Nichtwörter" und "Gedächtnis für Wörter" erfasst.

Fried et al. (2009) richten das Sprachkompetenzmodell von Delfin 4 am Konstrukt der "Akademischen Sprache" aus. Diese unterscheiden sie von der "Konversationssprache", die für die Kommunikation in vertrauten Situationen hinreichend ist. Sie führen aus, dass sich Kinder, deren Akademische Sprache gut ausgeprägt ist, über vertraute Interaktionssituationen hinaus "kontextfrei und präzise" ausdrücken können, sie verfügen über einen "zum Teil schon abstrakten Wortschatz, sowie differenzierte morpho-syntaktische Strukturen" (Fried et al., 2009, S. 15). Das Modell teilt sich auf in eine Wissensdimension, eine Könnensdimension, eine Bewusstheitsdimension und eine Aufgabendimension und bezieht sich in diesen eher auf (teilweise unbewusstes) Wissen als auf die grundlegenden kognitiven Fähigkeiten.

Delfin 4 und SETK 3-5 erheben den Anspruch, Sprachkompetenz von Kindern allgemein zu erfassen, dennoch unterscheiden sich ihre Schwerpunkte in der Konzeption dieser Kompetenz. Das Kompetenzmodell des SETK 3-5 bezieht sich eher auf grundlegende Fertigkeiten, insbesondere kognitive Fähigkeiten, die den ungestörten Spracherwerb erst möglich machen. Es beschreibt also die Voraussetzungen, die für einen weiteren ungestörten Spracherwerb als nötig erachtet werden, wie zum Beispiel eine ausreichende Wahrnehmung phonologischer Unterschiede. Das Kompetenzmodell von Delfin 4 dagegen bezieht sich eher auf den Ist-Zustand in der Sprachentwicklung und versteht sich als Test von vorhandenem (auch unbewusstem) Wissen, nicht der Prozesse und Fähigkeiten, die zum Aufbau dieses Wissens nötig sind. Daneben enthält das Sprachkompetenzkonzept von Fried et al. (2009), das diese als "Akademische Sprache" bezeichnen, die Anforderung, dass Kinder in der Kindertagesstätte in der Lage sind, sich auf fremde Personen einzulassen und mit diesen in ungewohnten Kontexten zu sprechen. Es ist plausibel anzunehmen, dass diese Komponente, die vielleicht als "Reife" oder "Selbstbewusstsein" beschrieben werden kann, im Zusammenhang mit einer erfolgreichen

Teilnahme am Schulalltag steht. Problematisch ist jedoch ihre Deutung als Teilkomponente der Sprachkompetenz im engeren Sinne.

Auffällig ist, dass der SETK 3-5 und Delfin 4, obwohl sie grundsätzlich unterschiedliche Sprachkompetenzmodelle zu Grunde legen, in drei von vier Aufgabentypen identisch sind (Nachsprechen von Sätzen, Morphologische Regelbildung, Nachsprechen von Kunstwörtern). Dies stellt den tatsächlichen Bezug zwischen Aufgabentypen und valide gemessenen Kompetenzkonstrukten für mindestens eines der beiden Verfahren in Frage. Insbesondere werden in der Konzeption von Fried et al. (2009) Aspekte der Sprachverarbeitung nicht thematisiert. Dennoch wird mit der Aufgabe "Sätze nachsprechen" ein Aufgabentyp gewählt, in dem die gemessene Leistung von der Automatisierung der Sprachverarbeitung beeinflusst wird.

Kany und Schöler (2010, S. 25) beschreiben in Übereinstimmung mit dem Ansatz des SETK 3-5 den Aufbau von Sprachkompetenz als den Aufbau prozeduraler Fertigkeiten: "Vorrangig sind jeweils Verstehens- und Produktionsfertigkeiten aufzubauen, die uns am Ende als automatisierte Routinen zur Verfügung stehen. Man stelle sich vor, man müsste beim Sprechen darüber nachdenken, wie man Sprachlaute artikuliert, Silben und Wörter hintereinander fügt und letztlich Äußerungen produziert". Kany und Schöler beziehen sich hier explizit auf die Automatisierung von Prozessen als Komponente des sprachlichen Könnens. Jedoch kontrastieren sie "Automatisierung" mit dem "bewussten Nachdenken" über Sprachprozesse. Dies entspricht jedoch nicht dem Begriff der Automatisierung, wie er in Modellen des Arbeitsgedächtnisses beschrieben wird (einen Überblick geben Moors & De Houwer, 2006). Als automatisiert werden zumeist Prozesse beschrieben, die unter anderem ohne die Beanspruchung von Kontrollprozessen ablaufen und damit ohne deren Kapazitäten zu belasten. Im obigen Zitat von Kany und Schöler heißt Automatisierung nicht unbedingt, dass diese Prozesse keine Ressourcen kontrollierter Verarbeitung beanspruchen, sondern lediglich, dass sie nicht durch bewusstes Nachdenken gesteuert werden müssen. Diese beiden Aussagen sind jedoch nicht identisch und ihre Unterscheidung ist zentral für die weitere Argumentation. Daneben verdeutlicht Kany und Schölers Aussage, wie unterschiedlich der Begriff Automatisierung verwendet wird. Im Folgenden soll daher das Konzept von Automatisierung erläutert und diskutiert werden, das in dieser Arbeit verwendet wird.

2.2 Automatisierung und Kontrolle

Die Eigenschaften automatisierter Prozesse werden häufig in Abgrenzung von kontrollierten Prozessen beschrieben (vgl. Rummer, 1996). Die sogenannte Dualitätsannahme (vgl. Neumann, 1992) besagt, dass alle Informationsverarbeitungsprozesse entweder als kontrollierte oder als automatisierte Prozesse beschrieben werden können. Automatisierte Prozesse sind schnell und robust, das heißt, sie können nicht (z.B. durch die Anforderungen einer Zweitaufgabe) gestört werden. Sie sind nicht intentional, nicht bewusst und unterliegen keiner Kapazitätsbegrenzung. Eine Übersicht über die Eigenschaften automatisierter und kontrollierter Prozesse innerhalb der Dualitätsannahme gibt Rummer (1996, s. Tabelle 1).

Tabelle 1: Die Abgrenzung automatischer und kontrollierter Informationsverarbeitungsprozesse (übernommen aus Rummer, 1996).

	kontrollierte Prozesse	automatisierte Prozesse
kapazitative Begrenzung	+	-
Intentionalität	+	-
Bewusstheit der Repräsentation	+	-
Geschwindigkeit und Robustheit	-	+

Unter der Dualitätsannahme ist ein Prozess entweder kontrolliert *oder* automatisiert, Misch- oder Zwischenzustände können unter dieser Annahme nicht konzeptualisiert werden. Da Misch- und Zwischenzustände jedoch in zahlreichen Experimenten beobachtet wurden (einen Überblick gibt Rummer, 1996), stellt Neumann (1992) der Dualitätsannahme ein Konzept entgegen, das diese Misch- und Zwischenzustände berücksichtigt. Rummer (1996, S. 37) greift diese Annahme auf: "Bei automatischen und kontrollierten Prozessen handelt es sich danach nicht um zwei Klassen mentaler Prozesse, sondern um zwei Pole eines (mehrdimensionalen) Konstrukts".

Bezieht man diese beiden Annahmen, die Dualitätsannahme und die Annahme zweier Pole Automatisierung und Kontrolle, auf die Verarbeitung einer Zweitsprache, so wird deutlich, dass von diesen beiden nur Letztere in der Lage ist, die Entwicklung

während des Erwerbs einer L2 zu beschreiben. Nimmt man mit der Dualitätsannahme an, dass eine L2 entweder kontrolliert oder automatisiert verarbeitet wird, so können nur der Anfangszustand des Spracherwerbs, nämlich das kontrollierte Anneinanderreihen von Wörtern gemäß einer bewusst angewandten syntaktischen Struktur und der Zustand des Muttersprachlers (der beim L2-Erwerb selten erreicht wird), dessen Verarbeitung gänzlich automatisiert abläuft, beschrieben werden. Diese Konzeption passt zu der Beschreibung von Kany und Schöler (2010), die die automatisierte Verarbeitung mit der vollständig bewussten kontrastieren (s. oben).

Die Dualitätsannahme erfasst aber nicht die Zwischenzustände, in denen sich weiter fortgeschrittene Lerner einer Zweitsprache befinden: Zumindest teilweise wird die L2 fließend gesprochen, die subjektive Anstrengung ist jedoch höher als bei einer Interaktion in der Muttersprache. Dennoch sind spontane Äußerungen in der L2 möglich, ohne dass z.B. über die syntaktische Struktur bewusst nachgedacht wird. Ebenso wenig erfasst wird der Zustand des hochkompetenten Sprechers einer L2, dessen Sprachverarbeitung automatisiert und ohne subjektiv erlebte Anstrengung abläuft, sich aber dennoch von der der Muttersprachler in ihrem Automatisierungsgrad unterscheidet.

2.3 *Grade der Automatisierung von Sprachverarbeitung*

Es ist also plausibel anzunehmen, dass ein Sprecher einer L1 und ein hochkompetenter Sprecher einer L2 im Grad der Automatisierung voneinander abweichen können, aber die Sprachverarbeitung dieses L2-Sprechers dennoch als automatisiert betrachtet werden muss. Treten diese Automatisierungsunterschiede in Abwesenheit von Performanzunterschieden in alltäglichen, bildungsrelevanten und der überwiegenden Anzahl experimenteller Testaufgaben auf, so sollten sie nicht als Unterschiede in der Sprachkompetenz gewertet werden. Unterschiede in der Sprachkompetenz sollten für das Sprachverhalten und die Teilhabe an einer Sprachgemeinschaft *relevante* Unterschiede abbilden.

Es ist nicht das Ziel dieser Arbeit, ein umfassendes Konzept von Sprachkompetenz zu entwickeln. Die Diskussion der Modelle weiter oben sollte jedoch die Schwierigkeiten bei der Konstruktion eines solchen Kompetenzmodells verdeutlichen. Sprachtests sehen sich neben der Konstruktion eines anwendbaren Modells zusätzlich vor die Herausforderung gestellt, das jeweilige Konstrukt mit den zur Verfügung stehenden

Messmethoden (also Aufgabenstellungen) adäquat zu erfassen. Dass auch dieser Schritt Probleme mit sich bringt, wurde anhand der sich überschneidenden Testaufgaben des SETK 3-5 und Delfin 4 deutlich, die gänzlich verschiedene Konzepte von Sprachkompetenz mit fast identischen Aufgabentypen messen.

Hauptsächlich aber sollte dieses Kapitel die Konzepte hinter den Begriffen Sprachkompetenz und Automatisierung verdeutlichen, die den Annahmen in dieser Arbeit zugrunde liegen. Die zentrale Annahme ist dabei, dass Automatisierungsunterschiede, die sich auf einem insgesamt sehr hohen Niveau der Automatisierung abspielen, nicht als Unterschiede in der Sprachkompetenz gedeutet werden sollten.

Im nächsten Kapitel wird auf die Unterschiede in der Verarbeitung einer L1 und L2 eingegangen. Dabei stehen insbesondere Unterschiede in der Automatisierung zwischen einer L1 und einer L2 im Vordergrund, die ohne erkennbare Unterschiede in der sprachlichen Performanz auftreten. Dies soll deutlich machen, dass der Fall unterschiedlicher Automatisierung bei flüssiger Sprachverarbeitung nicht nur theoretisch denkbar ist, sondern auch empirisch nachweisbar. Dabei wird besonders die Rolle der phonologischen Verarbeitung betont, das diese frühzeitig im Spracherwerb ein hohes Maß an Automatisierung erfährt und daher Unterschiede in der Automatisierung beim frühen Spracherwerb auf der phonologischen Ebene zu vermuten sind. Im ersten Schritt soll jedoch dargelegt werden, was hier unter Mehrsprachigkeit verstanden wird und welche Formen von Mehrsprachigkeit unterschieden werden.

3 Mehrsprachigkeit

In diesem Kapitel soll anhand theoretischer Konzepte und empirischer Befunde belegt werden, dass Automatisierungsunterschiede in der Verarbeitung einer L2 auch bei Sprechern auf höchstem Niveau der Sprachkompetenz aufzeigbar sind. Umgekehrt soll argumentiert werden, dass es für Sprecher einer L2 mit einem frühen Erwerbsalter nicht ungewöhnlich ist, (nahezu) muttersprachliche Sprachkompetenz zu erreichen. Um die dafür nötigen Konzepte von Mehrsprachigkeit zu erläutern, wird zunächst auf die verschiedenen Ausprägungen von Mehrsprachigkeit und unterschiedliche Erwerbsverläufe mit ihren Konsequenzen für die Sprachkompetenz und die Automatisierung der Sprachverarbeitung eingegangen. Der Zusammenhang von Erwerbsalter und der maximal erreichbaren Sprachkompetenz wird diskutiert. Es werden außerdem Befunde berichtet, die die Annahme stützen, dass Automatisierungsunterschiede zwischen einer L1 und L2 in Abwesenheit von Performanzunterschieden zu finden sind. Eine besondere Bedeutung kommt dabei der Automatisierung der Verarbeitung auf der phonologischen Ebene zu, die daher ausführlicher vorgestellt wird.

Der Begriff Bilingualität oder Mehrsprachigkeit wird in verschiedenen Kontexten unterschiedlich verwendet. In allgemeineren Kontexten wird der Begriff "bilingual" strenger verwendet, nämlich nur für Personen, die mit zwei Sprachen aufgewachsen sind. In dieser Arbeit bezeichnen die Begriffe "bilingual" und "mehrsprachig" jedoch allgemeiner Personen, die sich in zwei oder mehr Sprachen ausdrücken können und diese regelmäßig verwenden, ungeachtet des erreichten Niveaus ihrer Sprachkompetenz und ihres Erwerbsalters. Personen, die mit zwei Sprachen aufgewachsen sind, sollen in Abgrenzung davon als "simultan bilingual" bezeichnet werden. Da die Begriffe "bilingual" und "mehrsprachig" sehr weit gefasst sind, verwende ich für den mich interessierenden Personenkreis die Bezeichnung "*Near-Natives*". Damit sind Personen gemeint, die ihrer L2 mit einem zeitlichen Abstand von mindestens zwei Jahren nach ihrer L1 ausgesetzt waren und diese L2 auf einem Niveau beherrschen, dass dem von Muttersprachlern gleichkommt. Dies bedeutet, dass sie ihre L2 akzentfrei in verschiedenen Kontexten fließend anwenden, ohne Regerverletzungen, die nicht auch unter Muttersprachlern üblich sind. In Interaktionen mit der muttersprachlichen Sprachgemeinde werden sie üblicherweise für Muttersprachler gehalten. *Near-Natives* zeichnen sich in der Mehrzahl der Fälle durch ein frühes Erwerbsalter (*Age of Acquisition* - AoA) und eine Immersion in die Sprachgemeinde

der L2 aus. Typischerweise sind sie als Kinder von Einwanderern in das Land ihrer L2 gekommen.

Beim frühen Erwerb zweier Sprachen ist der sukzessive vom simultanen Spracherwerb abzugrenzen (z.B. De Houwer 2009; siehe auch McLaughlin 1978, Yip & Matthews, 2007). De Houwer unterscheidet die Erwerbstypen *Bilingual First Language Acquisition* (BFLA) und *Early Second Language Acquisition* (ESLA).

Beim Erwerbstyp *Bilingual First Language Acquisition* (BFLA) sind die Kinder ab ihrem ersten Lebensmonat zwei Sprachen ungefähr gleichermaßen kontinuierlich ausgesetzt und erwerben damit beide Sprachen von Geburt an. De Houwer (2009) spricht in diesem Fall nicht von einer L1 und L2 sondern von einer Sprache A und einer Sprache Alpha, um das Gleichgewicht der Sprachen zu betonen. Andere Forscher unterscheiden jedoch auch beim simultanen Spracherwerb zwischen einer L1 und einer L2. Sebastián-Gallés, Escheverría und Bosch (2005) werfen außerdem die allgemeinere Frage auf, ob der simultane Erwerb zweier Sprachen tatsächlich völlig balanciert ablaufen kann. Die Bezeichnung *Near-Native* ist im Falle des BFLA nicht angebracht, da die Sprecher als Muttersprachler in beiden Sprachen betrachtet werden sollten⁶.

Setzt der Erwerb der L2 nach dem 18. Lebensmonat aber vor Vollendung des vierten Lebensjahres ein und findet dann kontinuierlich parallel zur L1 statt, so bezeichnet De Houwer (2009) diesen Erwerbstyp als *Early Second Language Acquisition* (ESLA). Auch bei diesem Erwerbstyp ist es sehr wahrscheinlich, dass die Kinder ihre beiden Sprachen auf muttersprachlichem Niveau beherrschen, in ihrer L2 entsprechen sie damit dem zuvor beschriebenen Konzept des *Near-Native*-Sprechers.

ESLA beschreibt die Situation der Mehrheit der Kinder mit Migrationshintergrund, die gewöhnlich ihre L1 im Elternhaus und ihre L2 im Kindergarten bzw. in der Kindertagesstätte erwerben. Für die Annahmen und Studien in dieser Arbeit ist es entscheidend, L1 und L2 eindeutig unterscheiden zu können, sie beziehen sich daher nicht auf simultan bilingual aufwachsende Kinder. Daher wird in dieser Arbeit nicht auf den Erwerbstyp BFLA, also den simultanen Erwerb zweier Sprachen, sondern ausschließlich auf ESLA, den frühen sukzessiven Erwerb, und teilweise den später einsetzenden L2-Erwerb, eingegangen. Dass und unter welchen Umständen auch bei einem L2-Erwerb, der

⁶ Dies ist der Fall auch wenn eine der beiden Sprachen in geringerem Maße automatisiert sein könnte.

nach dem Alter von 4 Jahren einsetzt, ein nahezu muttersprachliches Sprachniveau (*near-native*) erreicht werden kann, wird im nächsten Abschnitt im Zusammenhang mit der *Critical Period Hypothesis*, der Hypothese einer kritischen Erwerbsphase, diskutiert.

3.1 Erwerbsalter und maximal erreichbare Sprachfertigkeit

Eine Reihe von Autoren nimmt an, dass die Möglichkeit zum nahezu muttersprachlichen Erwerb einer Sprache durch ein Zeitfenster beschränkt ist, dass also der Erwerb vor einem bestimmten Alter beginnen muss, um noch zur (nahezu) muttersprachlichen Sprachkompetenz führen zu können (vgl. DeKeyser, 2000). Diese Annahme wird üblicherweise als *Critical Period Hypothesis* bezeichnet, also die Annahme einer kritischen Phase im Spracherwerb. Birdsong (2006) definiert in seinem Übersichtsartikel zur kritischen Phase im Spracherwerb das Erwerbsalter (*Age of Acquisition* - AoA) als den Zeitpunkt der Immersion in den Kontext der L2. Als typischen Fall nennt er die Immigration in ein Land, in dem die L2 überwiegend gesprochen wird. Davon unterscheidet er das *Age of Exposure*, das früher liegen mag, zum Beispiel durch Schulunterricht in der L2. Wie zahlreiche Studien belegen, korreliert AoA gewöhnlich negativ mit dem *Ultimate Attainment* (UA), der individuell maximal erreichbaren Sprachkompetenz eines L2-Sprechers (einen Überblick gibt Birdsong, 2005). Dabei können sich die kritischen Erwerbsphasen für verschiedene sprachliche Verarbeitungsebenen (Phonologie, Morphologie, Syntax) unterscheiden (vgl. DeKeyser, 2000). Daneben wird das *Ultimate Attainment*, allerdings in geringerem Maße, von interindividuell abweichenden Faktoren wie der besonderen Fähigkeit zum L2-Erwerb, der Motivation und der psychosozialen Integration in die L2-Gemeinschaft beeinflusst.

Bei näherer Betrachtung muss der Zusammenhang zwischen AoA und UA jedoch differenzierter gesehen werden. Johnson und Newport (1989) zeigen für eine Gruppe mit Chinesisch oder Koreanisch als L1 und Englisch als L2, dass die Varianz in den Sprachfähigkeiten zwischen späten Lernern ($AoA > 16$) größer ist als zwischen frühen Lernern ($AoA \leq 16$), während über die Gruppen hinweg eine robuste negative Korrelation zwischen AoA und UA feststellbar ist. Diese negative Korrelation kann durch die Annahmen der *Critical Period Hypothesis*, die die Unterschiede durch Reifungsprozesse begründet sieht, erklärt werden. Die unterschiedlichen Varianzen deuten die Autoren als Hinweise darauf, dass bei einem späteren Spracherwerb ($AoA > 16$) interindividuelle Unterschiede stärker Einfluss auf den Spracherwerb nehmen, als bei einem frühen L2-

Erwerb ($AoA \leq 16$). Eine alternative Erklärung bietet das konkurrierende Modell von Bley-Vroman (1988), das in der *Fundamental-Difference-Hypothese* einen fundamentalen Unterschied in den Erwerbs- bzw. Lernmechanismen von Kindern und Erwachsenen annimmt. Dieses Modell besagt, dass Kinder ihre L2 durch die gleichen impliziten Mechanismen erwerben, die auch den Erstspracherwerb ermöglichen, während bei Erwachsenen diese Mechanismen nicht mehr zur Verfügung stehen und reflektierte Lernstrategien überwiegen. Erwachsene haben demnach Vorteile bei der Geschwindigkeit des Lernens einer L2, Kinder sind im Vorteil beim *Ultimate Attainment* (vgl. DeKeyser, 2000). DeKeyser (2000) replizierte den für die Gruppe der Zweitsprachler des Englischen und Chinesisch- bzw. Koreanisch-Muttersprachler von Johnson und Newport (1989) berichteten Befund mit Gruppen von Ungarisch-Englisch bilingualen Sprechern mit frühem ($AoA < 16$) und spätem ($AoA \geq 16$) Zweitspracherwerb. Dabei fand er eine robuste negative Korrelation von AoA und Sprachfähigkeit in der L2 über beide Gruppen hinweg, aber keine Korrelationen innerhalb der beiden Gruppen (vgl. DeKeyser, 2000, S. 510). Diese Befunde sind zwar im Einklang mit den Hypothesen von Johnson und Newport und von Bley-Vroman, demonstrieren aber auch den Einfluss weiterer Faktoren, die innerhalb der Gruppen wirken. Auch erreichen einige der Teilnehmer mit spätem AoA ein *Ultimate Attainment*, dass dem der Gruppe der frühen Lerner vergleichbar ist. Birdsong (2006) berichtet von Studien, in denen L2 Sprecher mit einem AoA von mehr als 11 Jahren ein muttersprachliches Niveau im Umgang mit ihrer L2 erreichten. Bialystok (1997) findet gar eine positive Korrelation zwischen AoA und Sprachfähigkeit in der L2, was der *Critical Period Hypothesis* zu widersprechen scheint. DeKeyser (2000) argumentiert jedoch, dass diese positive Korrelation durch überlegene Lernstrategien der älteren Lerner zustande kommt, die die L2 schneller erwerben. Da die Versuchspersonen von Bialystok (1997) vermutlich ihr *Ultimate Attainment* noch nicht erreicht haben, können die Befunde nicht als Belege gegen die *Critical Period Hypothesis* gewertet werden (vgl. DeKeyser, 2000). Dennoch demonstrieren sie, dass ein hohes Maß an Sprachkompetenz auch bei spätem Spracherwerb erreicht werden kann. Guillelmon und Grosjean (2001) finden in einer Benennungsaufgabe mit frühen ($AoA < 12$ Jahre) und späten ($AoA \geq 12$ Jahre) Englisch-Französisch Bilingualen Fazilitations- und Inhibitionseffekte durch kongruente und inkongruente Markierungen des grammatischen Geschlechts, allerdings nur für die Gruppe der frühen Bilingualen, was auf eine Verarbeitung morphosyntaktischer Strukturen in der L2 bei den frühen Bilingualen hindeutet, die der von Muttersprachlern entspricht. All diese Studien verwendeten behaviorale Daten, also Testergebnisse für verschiedene Komponenten sprachlicher Fähigkeiten (u.a. Syntax, Morpho-Syntax, Lexikon). Insgesamt

zieht Birdsong (2006) den Schluss, dass der Zusammenhang zwischen AoA und maximal erreichbarer Sprachkompetenz in einer L2 durch die sich mit steigendem Alter verändernden allgemeinen kognitiven Prozesse und Strukturen (wie z.B. Verarbeitungsgeschwindigkeit und Arbeitsgedächtnis) bestimmt wird. Die Effekte erklärt er daher weniger als kritische Phase im Sinne von abgeschlossenen Reifeprozessen mit einem fixen Endpunkt, sondern als kontinuierliche Entwicklung, die den Erwerb bzw. das Erlernen einer Sprache mit steigendem Alter verändert und ein muttersprachliches Level als *Ultimate Attainment* schwieriger und unwahrscheinlicher erreichen lässt, jedoch nicht gänzlich ausschließt.

Aus diesen Studien lässt sich schließen, dass es zumindest für L2-Sprecher mit frühem und mittlerem AoA (< 16) durchaus möglich ist, auch nach einem AoA von vier Jahren ein muttersprachliches Niveau im Sprachgebrauch, in Sprachtests und in experimentellen Erhebungen zu erreichen. Gleichzeitig, so soll im folgenden Unterkapitel demonstriert werden, sind auch bei hochkompetenten L2-Sprechern mit sehr frühem Erwerbsalter (< 3) Verarbeitungsunterschiede im Vergleich zu Muttersprachlern feststellbar. Dies spricht für die weiter oben theoretisch diskutierte Trennung von Sprachkompetenz und Automatisierung.

3.2 Automatisierung

Dieses Kapitel zielt insgesamt darauf ab, zu zeigen, dass muttersprachliche Kompetenz in einer L2 gewöhnlich nicht mit muttersprachlichen Verarbeitungsprozessen einhergeht. Bisher wurde gezeigt, dass Sprecher einer L2 mit frühem Erwerbsalter gewöhnlich ein nahezu muttersprachliches Niveau der Sprachkompetenz erreichen, wenn sie der Sprache (zumeist als Umgebungssprache) verstärkt ausgesetzt sind. Um nun zu argumentieren, dass muttersprachliche Kompetenz (das gezeigte sprachliche Verhalten) in einer L2 nicht zwingend an muttersprachliche Verarbeitung (die internen Prozesse, die diesem Verhalten zugrunde liegen) geknüpft ist, werden im Folgenden Belege für Verarbeitungsunterschiede zwischen Muttersprachlern und *Near-Natives* berichtet und diskutiert. Da unter diesen Belegen den Studien mit neuropsychologischer Methodik eine zentrale Rolle zukommt, wird auch auf die verwendeten Methoden in aller Kürze eingegangen.

Bei der Untersuchung von Unterschieden zwischen der Verarbeitung einer L1 und einer L2 werden zum Teil neuropsychologische Maße verwendet. In den hier aufgeführten Studien kommt unter anderem die Methode des *functional magnetic resonance imaging* (fMRI) zum Einsatz, die mittels eines Magnetresonanztomographen den Blutfluss im Gehirn während bestimmter Verarbeitungsschritte darstellen kann. Eine höhere Durchblutung bestimmter Hirnregionen wird als höhere Aktivität in diesen interpretiert. Stowe und Sabourin (2005) bezeichnen diese Methode als "wo"-Methode, da sie die von erhöhter Aktivität betroffenen Hirnareale abbilden kann, aber nicht den genauen zeitlichen Verlauf des Aktivitätsflusses. Im Kontrast dazu bezeichnen Stowe und Sabourin (2005) die Messung der Hirnaktivität mittels Elektroden als eine der "wann"-Methoden. Hierbei wird die Hirnaktivität mittels eines Elektro-Enzephalographen und Elektroden an der Kopfhaut registriert. Hierbei interessieren insbesondere die ereigniskorrelierten Potentiale (EKPs), positiv oder negativ ausfallende Ausschläge in der Hirnaktivität zu bestimmten Zeitpunkten relativ zu einem Ereignis. Da die Hirnaktivität zu keinem gegebenen Moment nur einen Prozess wiedergibt, sind nachträgliche Berechnungen zur Isolierung der interessierenden Aktivität nötig. Positive Ausschläge werden dabei jeweils mit einem P und der Angabe des zeitlichen Abstands von dem interessierenden Ereignis (in Millisekunden) gekennzeichnet, negative Ausschläge werden analog mit N gekennzeichnet. Als Beispiele für gut untersuchte EKPs möchte ich hier die beim Sprachverstehen häufig mit der semantischen Integration in Verbindung gebrachte N400, ein negativer Ausschlag 400 ms nach der Präsentation eines Wortes, und die P600, ein positiver Ausschlag, der 600 ms nach der Präsentation einer syntaktischen Verletzung zu beobachten ist, anführen (vgl. Kutas & Hillyard, 1980, 1983; Osterhout & Holcomb, 1992).

In experimentellen Studien, die nicht auf die oben beschriebenen Methoden zurückgreifen, werden die Prozesse beim Sprachverstehen häufig durch Lesestudien untersucht. Dabei werden unter anderem *Self-paced*-Lesestudien und *Eye-tracking*-Untersuchungen durchgeführt. Beim selbstgesteuerten (*self-paced*) Lesen werden Wörter einzeln in der Bildschirmmitte präsentiert. Die Versuchsperson kann sich jeweils durch das Drücken einer Taste das nächste Wort anzeigen lassen. Dabei wird der zeitliche Abstand zwischen dem mehrmaligen Drücken der Taste gemessen. Über die Lesegeschwindigkeit der Wörter werden Rückschlüsse auf die Verarbeitung des Satzes gezogen. Bei dieser Methode ist es für die Versuchsperson (im Kontrast zum normalen Leseprozess) nicht möglich, im Satz zurückzuspringen, um ein Wort noch einmal zu lesen. Daher kommen in den letzten Jahren vermehrt *Eye-Tracking*-Verfahren zum Einsatz. Bei diesen wird eine

Kamera auf das Auge des Probanden gerichtet, die die Bewegung einer seiner Pupillen beim Lesen registriert und aufzeichnet. So kann der zeitliche und räumliche Verlauf des Leseprozesses registriert werden, was weitere Rückschlüsse auf die zu Grunde liegenden Prozesse erlaubt.

Die Studien, die im Folgenden berichtet werden, beziehen sich ausschließlich auf *Near-Natives*. Automatisierungsunterschiede zwischen diesen und Muttersprachlern sind besonders aussagekräftig aufgrund des hohen Sprachniveaus der *Near-Natives*. Trotz ihrer hohen Sprachkompetenz unterscheiden sie sich im Grad der Automatisierung der Sprachverarbeitung von Muttersprachlern. Im Falle weniger kompetenter L2-Sprecher wäre dieser Unterschied trivial, würde nicht überraschen und müsste natürlich als zusammenhängend mit der Sprachkompetenz betrachtet werden, bei *Near-Natives* dagegen gelingt es, Automatisierungsunterschiede in Abwesenheit von anderweitig beobachtbaren Kompetenzunterschieden aufzuzeigen. Um die Ähnlichkeit von Muttersprachlern und *Near-Natives* in ihrer Performanz noch einmal zu bekräftigen, werden die Ergebnisse zweier experimenteller Studien ohne den Einsatz neuropsychologischer Verfahren berichtet, in denen kein Unterschied zwischen Muttersprachlern und *Near-Natives* gefunden wurde. Favreau und Segalowitz (1983, zitiert nach Segalowitz & Hulstijn, 2005) stellten in einer Studie zur Lesegeschwindigkeit und zum lexikalischen Priming keine Unterschiede zwischen Muttersprachlern und hochkompetenten Sprechern einer L2 fest. Frenck-Mestre (2002) fand in verschiedenen *Eye-Tracking*-Studien zur syntaktischen Verarbeitung keine Verarbeitungsunterschiede zwischen Muttersprachlern und hochkompetenten Zweitsprachlern.

Dies bekräftigt die Annahme, dass *Near-Natives* weder durch ihre Leistung in Sprachtests (mit Ausnahme der Satzwiedergabe), noch durch die Mehrheit experimenteller Designs von Muttersprachlern zu unterscheiden sind. In Messungen mit neuropsychologischer Methodik (z.B. fMRI und EEG) hingegen werden regelmäßig Unterschiede in der Hirnaktivität zwischen Sprechern einer L1 und einer L2 während sprachbezogener Aufgabenstellungen gefunden.

In einer EKP-Studie zur Verarbeitung von syntaktischen Strukturen bei Muttersprachlern und *Near-Natives* fanden Clahsen und Felser (2006) Verarbeitungsunterschiede zwischen den beiden Gruppen bei komplexen Strukturen, jedoch keine Unterschiede bei einfachen syntaktischen Strukturen. Andere Autoren fanden

mit neuropsychologischen Methoden Hinweise auf eine erhöhte Belastung der Aufmerksamkeit beim Verarbeiten einer L2. So berichten Stowe und Sabourin (2005) in einer fMRI-Studie von einer erhöhten Hirnaktivität bei hochkompetenten L2-Sprechern beim Verarbeiten ihrer L2 im Vergleich zu Muttersprachlern. Die Verarbeitung fand jedoch in den gleichen Hirnarealen statt, die auch bei der Verarbeitung einer L1 aktiv sind. Ähnliche Befunde berichten Perani, Abutalebi, Paulescu Brambati, Scifo, Cappa und Fazio (2003) in einer weiteren fMRI-Studie für Spanisch-Katalanisch bilinguale Sprecher, die ab dem Alter von drei Jahren ihren beiden Sprachen gleichermaßen ausgesetzt waren. Sie stellten eine erhöhte Hirnaktivität bei Sprechern einer L2 im Vergleich zu Sprechern einer L1 während einer lexikalischen Produktionsaufgabe fest. Stowe und Sabourin (2005) beziehen sich außerdem auf eine Studie von Weber-Fox und Neville (1996). Diese fanden in einer EKP-Studie Verarbeitungsunterschiede während einer morpho-syntaktischen und einer lexikalisch-konzeptuellen Beurteilungsaufgabe zwischen bilingualen und monolingualen Sprechern. Dabei unterschied sich sogar das Aktivierungsmuster der Gruppe von L2-Sprechern mit dem frühesten AOA (1-3 Jahre) vom Aktivierungsmuster der Muttersprachler. Stowe und Sabourin (2005, S. 179) schließen daraus, dass "even when an L2 is learned very early in life, different neural resources may act for its processing, which may not be associated to a worse language performance at the behavioral level. These cerebral differences may be particularly sensitive to effortful language tasks". Es soll noch einmal betont werden, dass sich die Verarbeitungsunterschiede in diesen Studien nicht in der Performanz der *Near-Natives* in den experimentellen Aufgabenstellungen widerspiegeln, diese ist auf muttersprachlichem Niveau. Die Unterschiede werden fast ausschließlich durch den Einsatz neuropsychologischer Methoden sichtbar. Die seltene Ausnahme (neben der Satz wiedergabe) eines experimentellen Verfahrens, das in der Lage ist, Verarbeitungsunterschiede von *Near-Natives* zumindest für die phonologische Verarbeitung ohne den Einsatz neuropsychologischer Methoden aufzuzeigen, bildet eine phonologische Diskriminierungsaufgabe (z.B. Sebastián-Gallés, Echeverría & Bosch, 2005). Auf dieses Verfahren und die damit gewonnenen Erkenntnisse zur phonologischen Verarbeitung einer L2 wird im folgenden Abschnitt eingegangen.

3.2.1 Frühe Automatisierung der phonologischen Verarbeitung

Da die Automatisierung der phonologischen Verarbeitung sehr früh erfolgt, sind Automatisierungsunterschiede zwischen *Near-Natives* und Muttersprachlern vermutlich überwiegend auf der phonologischen Verarbeitungsebene zu finden. Daher soll auf diese

Ebene der Sprachverarbeitung und die Befundlage zu Unterschieden zwischen Muttersprachlern und *Near-Natives* hier detaillierter eingegangen werden.

Hinweise auf eine frühe Automatisierung der phonologischen Verarbeitungsebene geben zahlreiche Studien zur phonologischen Diskriminierung mit sehr früh bis nahezu simultan bilingualen Sprechern (Sebastián-Gallés & Bosch, 2005; Sebastián-Gallés, Echeverría & Bosch, 2004; Bosch, Costa & Sebastián-Gallés, 2000; Navarra, Sebastián-Gallés & Soto-Faraco, 2005; Werker & Tees, 1984, 1999). Werker und Tees (1999, S. 509; s. auch 1984) beschreiben eine „transition from language-general to language-specific perceptual sensitivity that occurs during the first year of life“. Die Sensibilität für phonologische Unterschiede verringert sich in dem Maße, in dem sich die Phonemkategorien der Muttersprache verfestigen, bzw. die Verarbeitung der Laute als Phoneme automatisiert wird. Phonemkategorien bilden sich durch Unterteilungen auf dem Kontinuum der artikulierbaren Laute. Die Kategorisierung führt zu Veränderungen in der Wahrnehmung der Laute: Laute, die zu unterschiedlichen Kategorien gehören, werden als unterschiedlicher wahrgenommen als Laute, die innerhalb einer Phonemkategorie liegen, auch wenn sie auf dem Kontinuum gleich weit voneinander entfernt sind. Dies geschieht bei simultan bilingual aufwachsenden Kleinkindern später als bei Kindern, die zunächst monolingual aufwachsen. Während es bei monolingualen Kleinkindern Hinweise auf eine Festigung der Phonemgrenzen im Alter von 10-12 Monaten gab, hatten sich die Phonemsysteme der beiden Sprachen von bilingualen Kleinkindern erst im Alter von 14-21 Monaten herausgebildet (vgl. Sebastián-Gallés & Bosch, 2005). Dies kann gezeigt werden, indem den Kindern Laute präsentiert werden und ihre Aufmerksamkeit diesen Lauten gegenüber registriert wird. Als Aufmerksamkeit wird dabei die Zuwendung zur Quelle des Lautes gewertet. Wird der gleiche Laut wiederholt präsentiert, so gewöhnen sich Kinder an diesen Reiz und schenken ihm keine Aufmerksamkeit mehr. Wird ein neuer Reiz präsentiert, so wenden sie ihre Aufmerksamkeit wieder in Richtung des neuen Reizes. Um Aufschluss über die Festigung der Phonemgrenzen zu gewinnen, werden nun entweder Laute präsentiert, die innerhalb der Phonemgrenzen des muttersprachlichen Phonemsystems liegen oder über diese hinaus gehen. Von gefestigten Phonemgrenzen ist auszugehen, wenn die Laute, die in einer Phonemkategorie liegen nicht mehr als unterschiedliche Reize wahrgenommen werden, also keine erneute Aufmerksamkeitsreaktion auf die Präsentation des zweiten Lautes folgt, wenn das Kinde zuvor den ersten Laut gehört hat.

Doch auch bei simultan mit zwei Sprachen aufgewachsenen erwachsenen Sprechern stellen Sebastián-Gallés, Echeverría und Bosch (2005) eine Dominanz einer der Sprachen fest. In ihrer Studie untersuchten sie simultan Spanisch-Katalanisch bzw. Katalanisch-Spanisch aufgewachsene Sprecher, die sich darin unterschieden, welche Sprache (Spanisch oder Katalanisch) ihre Mutter mit ihnen sprach. Die jeweils andere Sprache wurde vom Vater der Teilnehmer mit ihnen gesprochen. Alle Teilnehmer waren in einer bilingualen Umgebung aufgewachsen, besuchten bilinguale Schulen und sprachen beide Sprachen innerhalb ihrer Familie. Zum Zeitpunkt der Erhebung waren sie Studenten an einer Universität in Barcelona. In einem Test mit einer lexikalischen Entscheidungsaufgabe (Wort – Nichtwort) wurden die kritischen Nichtwörter des Katalanischen durch einen Vokalkontrast (/e - ε/) gebildet, der nur im Katalanischen, aber nicht im Spanischen existiert. Die Fehlerraten der simultan bilingualen Sprecher mit Spanisch als Sprache der Mutter waren signifikant höher als die der simultan bilingualen Sprecher mit katalanischsprachigen Müttern. Die simultan Spanisch-Katalanischen Probanden schnitten jedoch besser ab, als eine Vergleichsgruppe von Spanischen-Katalanischen Probanden, die ihre L2 Katalanisch früh (aber nicht simultan) erworben hatte. Die Autoren diskutieren zwei alternative Erklärungen für diese Befunde: „Whether this [...] can be accounted for in terms of a precocious interference between a (to be established) dominant L1 and a (less established) L2 language (note that this hypothesis implies the assumption that infants give preference to one language over the other, for which conflicting evidence is available [...]) or by the inherent dynamics of simultaneously establishing the categories of both languages, remains an open question“ (Sebastián-Gallés et al., 2005, S. 252). Im Rahmen dieser Arbeit spielt es keine Rolle, welche dieser Erklärungen sich in der zukünftigen Forschung als richtig herausstellen wird. Für die Fragestellung dieser Arbeit ist der Befund entscheidend, dass die Verarbeitung einer L2, auch bei simultanem oder frühem Spracherwerb, sich von der Verarbeitung einer L1 teilweise unterscheidet.

Unterschiede in der Wahrnehmung eines deutschen Vokalkontrasts fanden Rinker, Alku, Brosch und Kiefer (2010) in einer EKP-Studie. Sie untersuchten Kinder mit L1 Türkisch und L2 Deutsch im Alter von 5-6 Jahren in einem deutschen Kindergarten anhand der *Mismatch-Negativity* (MMN). Die *Mismatch-Negativity* bezeichnet ein ereigniskorreliertes Potential mit negativem Ausschlag, das dann beobachtet werden kann, wenn zwei nacheinander präsentierte Laute nicht übereinstimmen. Dabei zeigten die türkisch-deutsch bilingualen Kinder eine signifikant geringere Ausprägung der MMN bei

einem Vokalkontrast, der nur im Deutschen, aber nicht im Türkischen existiert, im Vergleich zu Kindern der gleichen Altersgruppe mit L1 Deutsch. Dies war der Fall, obwohl alle türkisch-deutschen Kinder in Deutschland geboren waren und seit mehreren Jahren einen deutschen Kindergarten besuchten. Die Unterschiede, die Rinker et al. (2010) aufzeigen konnten, sind ein weiterer Hinweis auf Unterschiede in der Verarbeitung von Sprache speziell auf der Ebene der phonologischen Verarbeitung. Diese entstehen vermutlich auch trotz eines sehr frühen L2-Spracherwerbs.

Die oben beschriebenen Studien können wie folgt zusammengefasst werden. Unterschiede in der Verarbeitung einer Erst- und Zweitsprache bei hochkompetenten L2-Sprechern müssen sich nicht zwangsläufig in Sprachtests und experimentellen Maßen zeigen, können aber gewöhnlich durch neuropsychologische Messungen zumindest für einzelne Ebenen der Sprachverarbeitung aufgedeckt werden. Für die phonologische Verarbeitung wurden in experimentellen Messungen, die sich auf einzelne Laute der L2 konzentrierten, Leistungsunterschiede zwischen zwei Gruppen simultan-bilingualer Probanden gefunden in Abhängigkeit davon, ob die Testsprache die Sprache der Mutter oder des Vaters war. Die Probanden in dieser zuletzt genannten Studie wuchsen simultan bilingual auf und sind damit als Muttersprachler in beiden Sprachgemeinschaften zu betrachten (vgl. De Houwer, 2009). Dennoch zeigten sich Teilnehmer, deren Mütter Katalanisch mit ihnen sprachen, in der phonologischen Verarbeitung von katalanischen Vokalkontrasten den Teilnehmern überlegen, deren Mütter Spanisch mit ihnen sprachen. Aus diesem Befund kann geschlossen werden, dass Automatisierungsunterschiede zwischen der Verarbeitung einer L1 und L2 in fast allen Fällen vorliegen. Diese Verarbeitungsunterschiede führen jedoch nicht zwingend zu Unterschieden in der Sprachkompetenz, gemessen in experimentellen Erhebungen oder Sprachtests. Bei *Near-Natives* werden sie von den Sprechern selbst und ihren Interaktionspartnern nicht wahrgenommen, *Near-Natives* sind insofern in dieser Hinsicht nicht von Muttersprachlern zu unterscheiden.

Neben der experimentellen Untersuchung von Vokalkontrasten ist die Satz wiedergabe womöglich eine weitere Möglichkeit, die Verarbeitungsunterschiede experimentell nachzuweisen. Durch die hohe Belastung des Arbeitsgedächtnisses beim wörtlichen Behalten von Sätzen wirkt sich der vorhandene Unterschied im Automatisierungsgrad der Sprachverarbeitung in dieser Aufgabe, im Gegensatz zu der überwiegenden Mehrheit der Aufgaben, auch auf die Leistung der *Near-Natives* aus, so

dass sie im Vergleich zu Muttersprachlern schlechter abschneiden. Da die Annahme der hohen Belastung des Arbeitsgedächtnisses beim Satzbehalten für diese Argumentation zentral ist, soll im Folgenden Kapitel auf die Prozesse eingegangen werden, die der Satzwiedergabe in einer L1 zugrunde liegen.

4 Satzbehalten

4.1 Anwendungsfelder der Satzwiedergabe

Das wörtliche Wiederholen von Sätzen hat sich in Sprachentwicklungstests für Muttersprachler bewährt. Kany und Schöler (2010, S. 127) bezeichnen es als "eine der besten Aufgabenstellungen, mit der Spracherwerbsstörungen erkannt werden können".⁷ Im SETK 3-5 wird diese Aufgabe im Untertest 2, "Satzgedächtnis" verwendet, in Delfin 4; Stufe 1 ist das "Sätze nachsprechen" eine von vier Aufgaben.

Die Kinder werden bei dieser Aufgabe angewiesen, zuvor auditiv präsentierte Sätze wörtlich zu wiederholen. Zumeist werden diese Sätze von der Versuchsleitung vorgelesen, manchmal auch mit Hilfe von Wiedergabegeräten präsentiert. Ein offensichtlicher Vorteil des Verfahrens ist, dass das Verstehen und die Produktion von Sprache auf sehr ökonomische Weise getestet werden, sowie darüber hinaus das implizite Wissen über syntaktische und morpho-syntaktische Regeln. Zunächst wird die Wiedergabe erheblich erleichtert, wenn der Satzinhalt bei der Rezeption verstanden wurde. Das Verstehen des Satzes, so könnte argumentiert werden, ist bereits aussagekräftig hinsichtlich der Sprachkompetenz, da es Wissen über Satzstruktur und einen ausreichenden Wortschatz voraussetzt. Zum anderen muss der Inhalt behalten werden; hier wird angenommen, dass Chunkingprozesse, also das Zusammenfassen von kleinen Informationseinheiten in größere sinnvolle Einheiten, die Behaltensleistung positiv beeinflussen. Diese Prozesse setzen wiederum eine gute Kenntnis der syntaktischen Struktur voraus. Daneben helfen syntaktische und morphologische Kenntnisse bei der Wiedergabe, da zum Beispiel nicht jedes Wort in seiner morphologischen Komplexität erinnert werden muss, wenn allgemeine Regeln des Satzbaus angewandt werden können. Neben diesen Verarbeitungskomponenten der Aufgabe können durch die Behaltenskomponente Rückschlüsse auf die Kapazität des phonologischen Speichers bzw. des verbalen Arbeitsgedächtnisses gezogen werden, welche bei muttersprachlichen Kindern wiederum in Zusammenhang mit der Entwicklung des Wortschatzes stehen (z.B. Baddeley, Gathercole & Papagno, 1998). Kany und Schöler (2010, S. 126) schließen aus den üblicherweise erzielten Testergebnissen und den Fehlerarten, die dabei auftreten, dass Kinder die Sätze "nicht einfach papageienhaft nachplappern, sondern diese verstehen und

⁷ Neben seinem Einsatz in Sprachstandserhebungen und Sprachentwicklungstests für Kinder wird dieser Aufgabentyp auch in anderen klinischen Kontexten verwendet (z.B. Meyers, Volkert & Diep, 2000).

die Satzaussage mit Hilfe von Satzstrukturen, die sie bereits beherrschen und in ihrem Gedächtnis gespeichert haben, rekonstruieren". All dies lässt die Satzwiedergabe als vielseitiges und geeignetes Maß der Sprachkompetenz erscheinen.

Sowohl im SETK 3-5 als auch in Delfin 4 werden neben den sinnvollen ("normalen") Sätzen auch sinnlose Sätze verwendet, also Sätze, die zwar gewöhnliche Wörter verwenden, diese aber so kombinieren, dass der Satz semantisch nicht sinnvoll gedeutet werden kann (z.B. "Die klatschenden Stühle singen einen Becher."). Die Sätze enthalten also mehrere semantische Anomalien. Semantische Anomalien und ihre Verarbeitung beim Satzverstehen sind unter anderem von Kutas und Federmeier (2000, s.a. Kutas & Hillyard, 1980, 1983; Osterhout & Holcomb, 1992) untersucht worden. Sie geben einen Überblick über die Forschung zur Verarbeitung semantischer Anomalien mittels ereigniskorrelierter Potentiale (EKP). EKPs sind Ausschläge in der Hirnaktivität, die durch Elektroden an der Kopfhaut gemessen werden können. Kutas und Federmeier (2000) präsentierten ihren Probanden Sätze, die zum Satzende hin jeweils ein im Kontext des Satzes plausibles oder nicht plausibles Wort enthielten. Währenddessen maßen sie die Hirnaktivität der Probanden mittels eines Elektro-Enzephalographen (EEG). Sie berichten, dass die bereits erwähnte N400 (ein negativer Ausschlag der Aktivität, der 400 ms nach der Rezeption eines Wortes gemessen wird und mit der semantischen Integration in Verbindung gebracht wird, vgl. z.B. Osterhout & Holcomb, 1992) unterschiedlich hoch ausfällt, je nachdem ob das Wort im gegebenen Kontext plausibel ist oder nicht. Kontexte können Wortlisten oder vorher gegebene Informationen sein, meist jedoch handelt es sich um den Kontext eines Satzes. Bei nicht plausiblen Wörtern fällt die N400 höher aus als bei Wörtern, die plausibel im gegebenen Kontext auftreten können. Der niedrigste Ausschlag der N400 wird gemessen, wenn das entsprechende Wort durch den Kontext voraktiviert wurde (vgl. Kutas & Federmeier, 2000).

Die Betrachtung der sinnlosen Sätze in den Sprachstandserhebungen müsste unter sorgfältiger Beachtung der Forschung zur Verarbeitung semantischer Anomalien geschehen. Den Einfluss der Verarbeitungsunterschiede zwischen den beiden Satzarten (sinnvoll und sinnlos) auf die Behaltensleistung zu modellieren und zu prüfen (immer unter Beachtung der möglicherweise unterschiedlichen Reaktionen auf semantische Anomalien in einer L1 und L2) würde den Rahmen dieser Arbeit sprengen. Daher beschränke ich meine Betrachtung der Satzwiedergabe auf die Wiedergabe sinnvoller, also semantisch plausibler Sätze.

Seltener als in Sprachstandserhebungen mit Kindern und in klinischen Kontexten wird die Satzwiedergabe zum Testen von gesunden Erwachsenen in einer L2 eingesetzt. Die Soziolinguisten Radloff (1991) und Diller und Jordan-Diller (2003) verwenden in Studien mit gesunden Erwachsenen das Nachsprechen von Sätzen als Aufgabe in Sprachscreenings in bilingualen Gemeinschaften zur Einschätzung des Grads an Bilingualität. Radloffs und Diller und Jordan-Dillers Studien unterscheiden sich jedoch von den hier diskutierten hinsichtlich ihrer Zielgruppe (ausschließlich mehrsprachige Personen, die aber ihre Zweitsprache nicht alle auf hohem Niveau beherrschen), des verwendeten Materials⁸ und der Fragestellungen (vgl. Radloff, 1991; Diller & Jordan-Diller, 2003). Die Studien von Radloff und Diller und Jordan-Diller werden hier nur aus Gründen der Vollständigkeit erwähnt.

Obwohl die Methode der Satzwiedergabe häufig zur Sprachstandsmessung eingesetzt wird und trotz ihrer zahlreichen Vorzüge ist es fraglich, ob tatsächlich die Sprachverarbeitung oder gar die allgemeine Sprachkompetenz für die Leistung beim Satzbehalten entscheidend sind. Arbeitsgedächtnisprozesse beeinflussen die Sprachverarbeitung maßgeblich (vgl. Rummer, 2003; Rummer, & Engelkamp 2011). Das Behalten eines Satzes ist durch die explizite Behaltensanforderung noch stärker von der Interaktion der (bei der Satzwiedergabe überwiegend parallel ablaufenden) Prozesse der Sprachverarbeitung und des Behaltens von Information beeinflusst. Bereits während des Lesens des Satzes müssen Inhalt und Struktur aktiv gehalten werden, später während der Reproduktion müssen Inhalt und Struktur bis zur nahezu vollständigen Wiedergabe aktiv gehalten werden. Aaronson und Scarborough (1976) finden in einer Lesestudie eine höhere kognitive Belastung beim Lesen eines Satzes, der danach wiedergegeben werden soll, im Vergleich zum Lesen eines Satzes ohne Behaltensaufgabe. Dies schlossen sie aus der unterschiedlichen Lesegeschwindigkeit, die ihre Probanden je nach Aufgabenstellung zeigten: Probanden, die instruiert wurden, den Satz nur zu verstehen, lasen schneller als Probanden, die instruiert wurden, den Satz später wörtlich wiederzugeben. Dies ist ein Hinweis darauf, dass sich die Verarbeitung eines Satzes unterscheidet, je nachdem, ob er verstanden oder behalten werden soll.

Die wörtliche Satzwiedergabe stellt eine hohe und ungewöhnliche Belastung für das Arbeitsgedächtnis dar. Nicht nur muss viel Information gespeichert bzw. aktiv gehalten

⁸ Radloff (1991) empfiehlt für die Aufgabe ausschließlich Sätze, die kurz genug sind, um von Muttersprachlern problemlos vollständig wiedergegeben zu werden. Dies ist nicht der Fall in den hier diskutierten Sprachstandserhebungen, in denen auch Muttersprachler meist nicht den vollständigen Satz reproduzieren.

werden, sondern es handelt sich teilweise auch um Satzoberflächeninformation (also die lexikalische, syntaktische, morpho-syntaktische und phonologische Umsetzung des Inhalts), deren Erinnern üblicherweise in der Kommunikation nicht relevant ist. Diese hohe und ungewöhnliche Belastung des Arbeitsgedächtnisses führt bei langen Sätzen zu einer Auslastung der Ressourcen des Arbeitsgedächtnisses über deren Grenzen hinaus, durch welche wiederum die Unterschiede in der Automatisierung der Sprachverarbeitung, die zwischen Sprechern einer L1 und L2 bestehen, auf die Leistung in der Satzwiedergabe wirken: Es stehen in einer L2 aufgrund der aufwändigeren Sprachverarbeitung weniger Ressourcen für die Aktivhaltung der Informationen zur Verfügung, die Leistung in der Satzwiedergabe wird schlechter.

Diese Argumentation basiert auf Modellen zum Arbeitsgedächtnis und zur Sprachverarbeitung. Im Folgenden sollen verschiedene Modelle des verbalen Arbeitsgedächtnisses und ihre Aussagen zum Satzbehalten vorgestellt werden (vgl. Just & Carpenter, 1992; Jefferies, Lambon Ralph & Baddeley, 2004; s. auch Caplan & Waters, 1999). Dabei kommt dem Rahmenmodell des Satzbehaltens von Rummer und Schweppe (2009, s.a. Schweppe 2006) eine besondere Rolle zu, da sich aus dessen Vorhersagen in Kombination mit den bereits diskutierten Studien zur Automatisierung einer L2 die zentralen Fragestellungen dieser Arbeit ergeben.

4.2 Einfluss zentraler exekutiver Prozesse auf das Satzbehalten

In diesem Abschnitt sollen die Befunde und theoretischen Vorstellungen zum Satzbehalten innerhalb einflussreicher Modelle des Arbeitsgedächtnisses dargelegt werden. Dabei soll auch, sofern vorhanden, auf die Annahmen zu kontrollierten und automatisierten Prozessen innerhalb der Modelle eingegangen werden.

Großen Einfluss auf die experimentelle Gedächtnisforschung hat seit vielen Jahren das Mehrkomponentenmodell des Arbeitsgedächtnisses von Baddeley und Hitch (1974; siehe auch Baddeley, 1986, 2000). Das Modell beinhaltet in seiner ursprünglichen Fassung drei Komponenten: die phonologische Schleife und die visuell-räumliche Komponente als Hilffsysteme sowie eine zentrale Exekutive, welche diese beiden steuert. Die zentrale Exekutive steuert dabei die kontrollierte Verarbeitung von Inhalten, ihre Kapazität ist beschränkt. In der aktualisierten Version des Modells (Baddeley, 2000) werden die drei

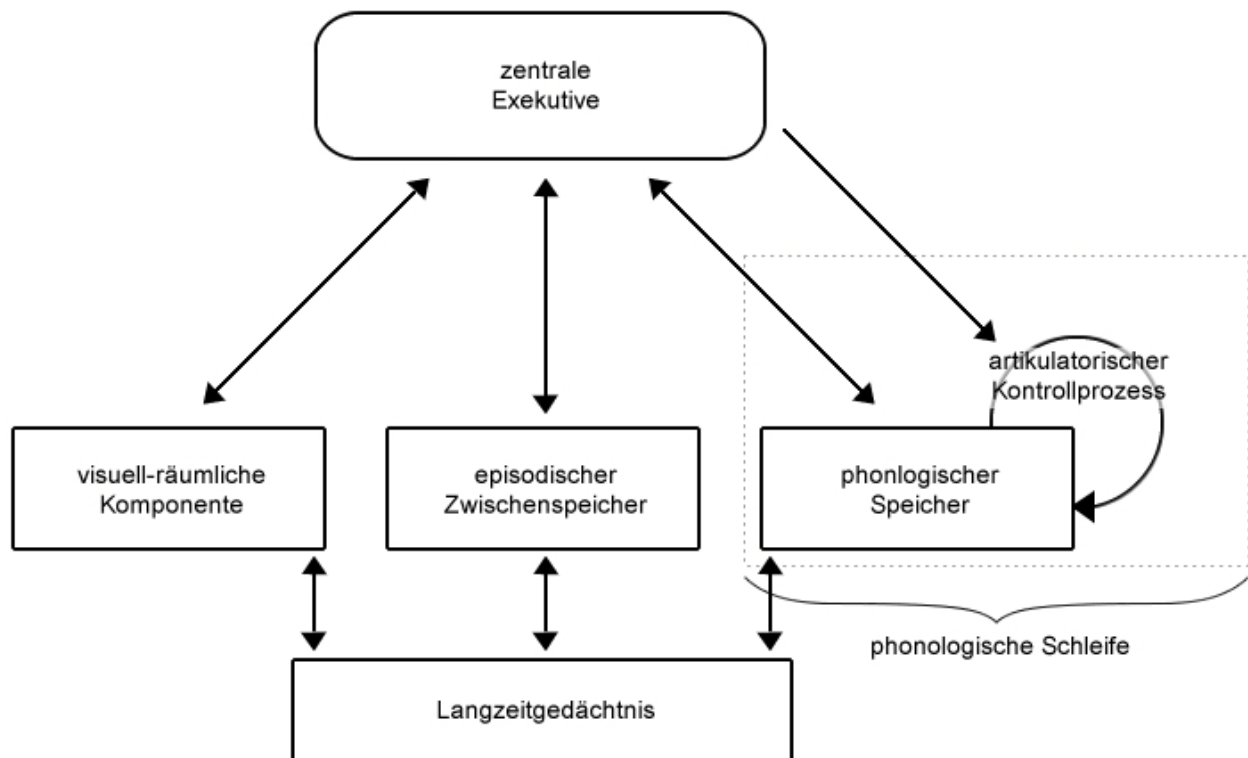


Abbildung 1: Schematische Darstellung des Mehrkomponentenmodells des Arbeitsgedächtnisses von Baddeley (2000).

Komponenten um Schnittstellen zwischen den Hilfsystemem und dem Langzeitgedächtnis (LZG), sowie den episodischen Zwischenspeicher ergänzt, der als Schnittstelle zwischen dem LZG und der zentralen Exekutive dient und den Transfer von Informationen vom Arbeitsgedächtnis ins LZG steuert (vgl. Abbildung 1). Der episodische Zwischenspeicher wird ebenfalls von der zentralen Exekutive gesteuert und benötigt daher Aufmerksamkeitsressourcen, also Ressourcen der kontrollierten Verarbeitung. Die phonologische Schleife besteht aus dem phonologischen Speicher und dem artikulatorischen Kontrollprozess. Informationen im phonologischen Speicher zerfallen sehr schnell, können aber durch Wiederholung (*Rehearsal*) kontrolliert aktiv gehalten werden. Das Mehrkomponentenmodell erklärt zahlreiche empirische Befunde (z.B. Effekt Phonologischer Ähnlichkeit, Artikulatorische Unterdrückung, *Irrelevant Speech Effect*) und gilt daher als gut belegt (vgl. z.B. Conrad & Hull, 1964; Murray, 1968; Baddeley, Thomson & Buchanan, 1975; Salamé & Baddeley, 1982). Der überwiegende Teil der Forschung innerhalb dieser Modellvorstellung verwendet sprachliches Material unterhalb der Satzebene, also Wörter und Laute. Die eigentlichen linguistischen Verarbeitungsprozesse (neben den Behaltensprozessen) werden als automatisiert angenommen (vgl. Jefferies, Lambon Ralph & Baddeley, 2004). In ihrer Studie zum Satzbehalten innerhalb des Mehrkomponentenmodells stellen Jefferies, Lambon Ralph und

Baddeley (2004, S. 623) jedoch einen Einfluss aufmerksamkeitskontrollierter Prozesse beim Satzbehalten fest: "Our results suggest that sentence recall reflects contributions from both automatic linguistic processes and attentionally limited working memory". Sie untersuchten anhand einer aufmerksamkeitsfordernden Zweitaufgabe die Anteile automatisierter und kontrollierter Prozesse beim Behalten kurzer Geschichten, bei Sätzen und bei Wortlisten. Dabei zeigte sich das Behalten von Sätzen als stärker beeinträchtigt durch die Zweitaufgabe als das Behalten von Wortlisten und Geschichten. Die Autoren schlossen daraus auf eine Beteiligung automatisierter und kontrollierter Prozesse am Satzbehalten.

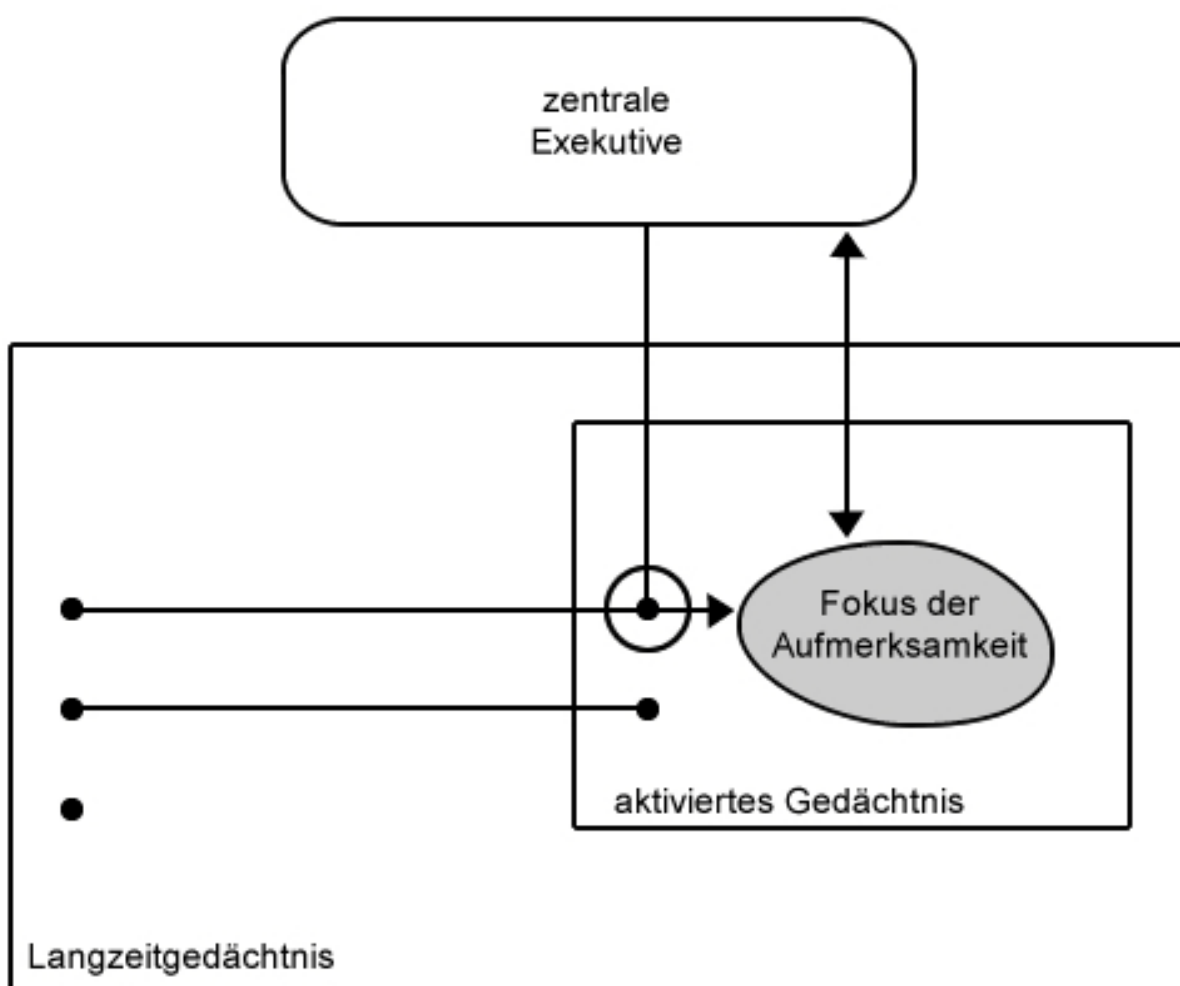


Abbildung 2: Schematische Darstellung des Modells des Arbeitsgedächtnisses als aktiviertes Langzeitgedächtnis (vgl. Cowan, 1988, 1993, 1999). Die einzelnen Punkte stellen Repräsentationen im Gedächtnis dar, die sich in unterschiedlichen Aktivierungszuständen befinden: nicht aktiviert, aktiviert und nicht im Fokus der Aufmerksamkeit, aktiviert und im Fokus der Aufmerksamkeit.

Neben dem Mehrkomponentenmodell gewinnen konkurrierende Konzeptionen des Arbeitsgedächtnisses in den letzten Jahren immer mehr an Bedeutung. Rummer und

Fiebach (2010) nennen für diese Entwicklung drei maßgebliche Gründe: die Häufung an empirischen Befunden, die das Mehrkomponentenmodell nicht erklären kann, die unzureichende Konzeption der visuell-räumlichen Komponente und die Widersprüche zwischen der neuronalen Implementierung des Modells von Baddeley mit Erkenntnissen aus der neurophysiologischen Forschung an nichtmenschlichen Primaten.

Unter den alternativen Konzeptionen des Arbeitsgedächtnisses (z.B. Anderson, 1993; Shiffrin 1993; Schneider & Detweiler, 1987, Engle, Kane & Tuholski, 1999; Oberauer, 2003) hat sich das *embedded process model* von Cowan (1988, 1993, 1999) in besonderer Weise etablieren können, da es sowohl Befunde innerhalb der Forschung zu Baddeley und Hitches (1974) Modell, als auch dem letzteren widersprechende Befunde erklären kann (vgl. Cowan, 1994, Cowan & Morey, 2007; Cowan, Saults & Brown, 2004). An die Stelle modularer und strukturell getrennter Komponenten, die über Schnittstellen interagieren, tritt bei Cowan die Konzeption des Arbeitsgedächtnisses als funktionaler Aktivierungszustand innerhalb des Langzeitgedächtnisses (vgl. Abbildung 2). Er beschreibt das Arbeitsgedächtnis als „cognitive processes that retain information in an unusually accessible state“ (Cowan, 1999, S. 62). Strukturell baut sein Modell auf Langzeitrepräsentationen und sensorischem Input auf, die durch Zuführung von Aktivierungsenergie aktiv gehalten werden. Die höchste Aktivierung haben Repräsentationen inne, die im Fokus der Aufmerksamkeit, also im Bewusstsein, sind.

Die Aufmerksamkeit, also der Mechanismus, der den Fluss der Aktivierungsenergie bestimmt, kann zum einen willentlich gesteuert werden und zum anderen unwillkürlich über sensorische Reize gesteuert werden. Repräsentationen im Arbeitsgedächtnis, die zum Beispiel kürzlich im Fokus der Aufmerksamkeit waren oder durch andere Prozesse eine Erhöhung der Aktivierung erfahren haben, sind auf einem höheren Aktivierungslevel im Vergleich zu anderen Repräsentationen im Langzeitgedächtnis. Dieser Zustand der höheren Aktivierung ist laut Cowan das, was das Arbeitsgedächtnis ausmacht. Die Aktivierung zerfällt über die Zeit. Werden die Repräsentationen (erneut) in den Fokus der Aufmerksamkeit gebracht, bevor die Restaktivierung zerfallen ist, so ist umso weniger zusätzliche Aktivierungsenergie nötig, je größer die noch vorhandene Restaktivierung ist. In diesem Zustand der Vor- bzw. Restaktivierung und im Fokus der Aufmerksamkeit sind diese Repräsentationen im Arbeitsgedächtnis, also leichter zugänglich als andere Repräsentationen. Anhand Cowans Modell lassen sich insbesondere Einflüsse von Langzeitrepräsentationen auf die

Arbeitsgedächtnisleistung erklären. Die Interaktion zwischen Arbeitsgedächtnis und Langzeitgedächtnis ist in diesem Modell klarer konzipiert und damit auch die Schnittstelle zwischen sprachlichen Repräsentationen (die ja eine Teilmenge der Langzeitgedächtnisinhalte bilden) und dem Arbeitsgedächtnis.

Cowan macht in seinem Modell keine direkten Aussagen zum Satzbehalten. Das Modell enthält jedoch die Aussage, dass Aktivierungsenergie und Aufmerksamkeit begrenzt sind. Auf je mehr Repräsentationen sie sich aufteilen, desto weniger Aktivierung kommt jeder einzelnen Repräsentation zu. Damit ist diese schwieriger verfügbar und kann eventuell nicht mehr erinnert werden. Wird Aufmerksamkeit zusätzlich für andere Prozesse benötigt (zum Beispiel die nicht vollständig automatisierte Sprachverarbeitung), so steht in diesem Modell weniger Aufmerksamkeit zur Aufrechterhaltung der Aktivierung der Repräsentationen zur Verfügung. Dies würde zu einem Verlust von im Arbeitsgedächtnis gespeicherten Informationen führen, da diese nicht mehr ausreichend aktiv und damit abrufbar sind. Gerät die Kapazität der Aufmerksamkeit an ihre Grenzen, so kann angenommen werden, dass umso weniger Repräsentationen verfügbar gehalten werden können, je mehr Steuerung und Aktivierungsenergie für zusätzliche Prozesse benötigt wird.

Wiederholt habe ich nun schon vom Zerfall oder vom Verlust von Information im Arbeitsgedächtnis gesprochen. Daher soll kurz dargelegt werden, welche Prozesse als ursächlich für diesen Verlust von Informationen angenommen werden. Oberauer und Kliegl (2010) fassen drei Klassen von Hypothesen zusammen, die Prozesse beschreiben, welche zum Verlust von Information im Arbeitsgedächtnis führen. Die erste ist die Spurenerfallhypothese (z.B. Baddeley & Hitch, 1974), die besagt, dass Information zerfällt, wenn sie nicht wiederholt wird. Diese findet sich in Baddeley und Hitchs Vorstellung der Notwendigkeit eines artikulatorischen Kontrollprozesses. Die zweite ist die Ressourcenhypothese, die eine limitierte Ressource (Aufmerksamkeit, Aktivierungsenergie) annimmt. Diese wird den betreffenden Repräsentationen zugeführt; je mehr Repräsentationen aktiv gehalten werden müssen bzw. je mehr Steuerung durch die Aufmerksamkeit erforderlich wird, desto weniger Aktivierung erhält jede einzelne. Je weniger Aktivierungsenergie verfügbar ist, um eine Repräsentation aktiv zu halten, desto schwieriger wird es, sie zu abzurufen. Die dritte Hypothese ist die Interferenzhypothese, welche besagt, dass Repräsentationen, die gleichzeitig im Arbeitsgedächtnis gehalten werden, einander gegenseitig stören. Da auch der Rehearsalprozess der Steuerung durch

die zentrale Exekutive bedarf und auch Interferenz durch Aufmerksamkeit gemindert werden kann, kann zusammenfassend gesagt werden, dass unter all diesen Hypothesen der Aufmerksamkeit als Prozess, der die Aktivierungsenergie steuern kann, eine entscheidende Rolle im Prozess zukommt.

Wie oben bereits angedeutet, besteht ein weiterer Unterschied zwischen dem Mehrkomponentenmodell und dem *embedded process model* darin, dass Letzteres die Konzeption einer Schnittstelle zwischen den Prozessen des Arbeitsgedächtnisses und dem Langzeitgedächtnis und damit auch den sprachlichen Repräsentationen vereinfacht. Für Cowan sind linguistische Repräsentationen im Arbeitsgedächtnis eine Teilmenge der Langzeitrepräsentationen, die durch Aktivierung in den Zustand Arbeitsgedächtnis gelangen. Die Schnittstelle zum Sprachsystem ist einfach zu modellieren, da Sprachverarbeitung ebenfalls von zahlreichen Forschern als Fluss von Aktivierungsenergie durch ein Netzwerk von Repräsentationen konzipiert wird (z.B. Dell & Reich, 1981; Dell, 1986; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Spivey & Tanenhaus, 1998; Engelkamp & Rummer, 1999; Friederici, 2002; Crocker & Pickering, 2000; Rumelhart & McClelland, 1986).

Andere Modelle verzichten gänzlich auf die Vorstellung des Arbeitsgedächtnisses als eigenem System, sondern betonen seine Natur als Schnittstelle bzw. Verknüpfung verschiedener Systeme. Laut Majerus (2010, S. 15) kann Arbeitsgedächtnis als "Systemverbund beschrieben werden, der von verschiedenen kognitiven Systemen und deren kontrollierter Verknüpfung abhängt". Er konzipiert das Kurzzeitgedächtnis als integratives Modell von Aufmerksamkeit, Sprachverarbeitung und Kurzzeitgedächtnis, wobei das Kurzzeitgedächtnis ein "Nebenprodukt der Ausrichtung selektiver Aufmerksamkeitskapazitäten auf Sprach- und Reihfolgenverarbeitungssystem während der Ausführung einer bestimmten Aufgabe" ist (Majerus, 2010, S. 15). Der Aufmerksamkeit kommen in Majerus' Modell die Funktion und Eigenschaften zu, die sie auch im *embedded process model* inne hat. Die Annahmen zum Satzbehalten, die sich aus Majerus' Modell ableiten lassen, sind analog zu denen, die für Cowans Modell oben bereits ausgeführt wurden.

Das Behalten von Sätzen wird maßgeblich von Prozessen des verbalen Arbeitsgedächtnisses und der Sprachverarbeitung und der Interaktion dieser Prozesse beeinflusst. In den letzten Abschnitten wurden die Aussagen verschiedener Modelle des

Arbeitsgedächtnisses und ihre Aussagen zum Satzbehalten diskutiert. Dabei zeigte sich, dass kontrollierte Prozesse für die Steuerung der Aktivierungsenergie und damit die Aufrechterhaltung von Informationen eine zentrale Rolle spielen.

4.3 Satzbehalten

Zahlreiche theoretische Überlegungen und empirische Befunde weisen darauf hin, dass Sprachverarbeitung auf Wort- und Satzebene ein mehrstufiger Prozess ist (z.B. Garrett, 1976, 1984; Levelt, 1993; Levelt, Schriefers, Vorberg, Meyer, Pechmann & Havinga, 1991; Roelofs 1992; , Levelt, Roelofs & Meyer, 1999; Dell & Reich, 1981; Dell, Chang & Griffin, 1999; Cutler, 1988; Caramazza, 1997; Butterworth, 1980; Fromkin, 1971; siehe auch Dell, Juliano & Govindjee, 1993). Sowohl bei der Produktion als auch bei der Rezeption wird eine Abfolge von Stufen der Sprachverarbeitung angenommen. So werden beim Prozess der Rezeption eines Satzes laut Friederici (z.B. 2002) und Rummer (z.B. 2003; s. auch Rummer & Engelkamp, 2011) mehrere Stufen von der phonologischen über die morpho-syntaktische, syntaktische und lexikalische Ebene durchlaufen. Am Ende des Prozesses steht der propositionale Gehalt des Satzes, also seine Botschaft, welche die Basis für (Re-)Aktionen, wie zum Beispiel Antworten und Handlungen, bildet (vgl. Abbildung 4). Während der (Re-)Produktion werden die gleichen Stufen in umgekehrter Reihenfolge durchlaufen, von der Botschaft bis zu den motorischen Plänen zur Realisierung der Sprachbotschaft und ihrer Ausführung. Diese grundlegende Konzeption der Sprachverarbeitung als mehrstufigem Prozess in Rezeption und Produktion liegt den Annahmen über das Satzbehalten zu Grunde, die im nächsten Abschnitt diskutiert werden.

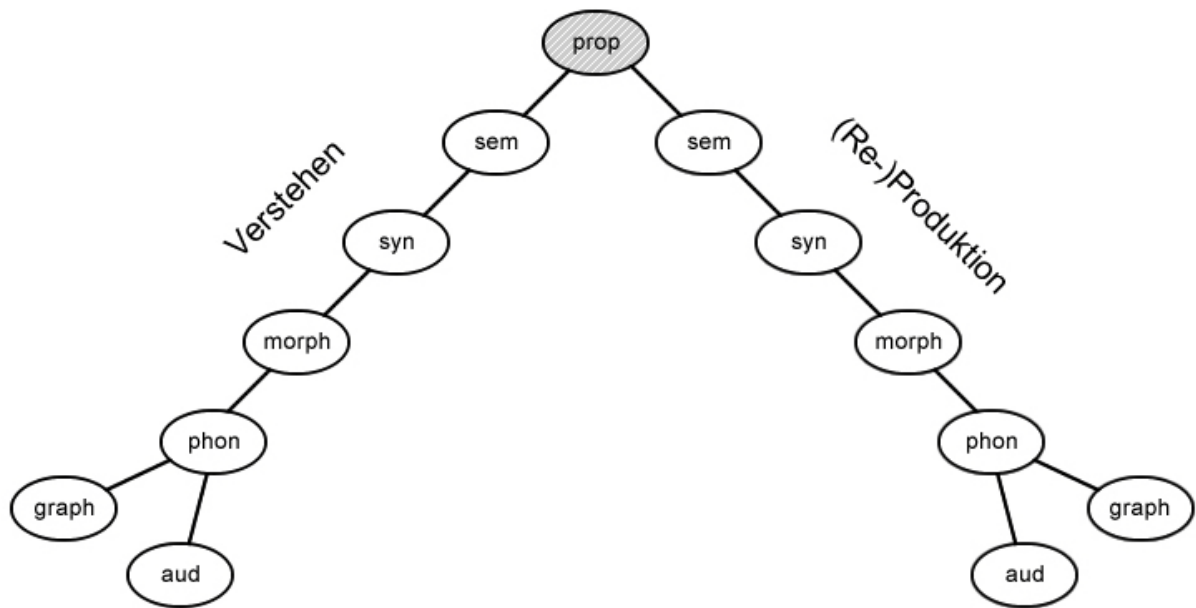


Abbildung 3: Schematische Darstellung der stufenweisen Verarbeitung von Sprache.

4.3.1 Satzbehalten: Sprachverarbeitung und Aufmerksamkeit

Um die für diese Arbeit zentralen Prozesse der Satzverarbeitung und des Behaltens zu verdeutlichen, werden in diesem Abschnitt Modelle des Satzbehaltens vorgestellt und ihr Zusammenhang mit der begrenzten Ressource Aufmerksamkeit diskutiert. Die Unterschiede zwischen der Satzverarbeitung zum reinen Verstehen und der Satzverarbeitung zum wörtlichen Behalten und empirische Hinweise auf diese Unterschiede werden dargelegt.

Im letzten Abschnitt wurde die Botschaft (auch: der propositionale Gehalt) des Satzes als Resultat der Satzverarbeitung beim Rezeptionsprozess genannt. Ebenso ist die Botschaft Ausgangspunkt der Satzverarbeitung im Produktionsprozess. Auf diese Annahme stützt sich die von Potter und Lombardi (1990) formulierte Regenerierungshypothese zum wörtlichen Satzbehalten (s. auch Lombardi & Potter, 1992). Potter und Lombardi nehmen an, dass Sätze bei der Wiedergabe nicht als Kette von Wörtern aus einem Speicher abgerufen werden, sondern dass sie während der Wiedergabe aus ihrem propositionalen Gehalt, also der Summe ihrer Aussagen, regeneriert werden. „What is directly represented in short-term memory is the sentence's conceptual message. In recall, normal mechanisms of sentence production are used to express this message, drawing on currently active lexical representations.“ (Lombardi & Potter, 1992, 713). Beim Verstehen werden die Lexeme der dargebotenen Wörter durch die Rezeption

aktiviert, aus diesen wird während der Rezeption die Botschaft extrahiert. Bei der Reproduktion wird aus dieser Botschaft der Satz regeneriert. Während dieses Regenerationsprozesses werden bevorzugt bereits aktivierte Lexeme verwendet. Dies resultiert in einer zumeist wörtlichen Wiedergabe des Satzes. Evidenz für diese Annahme finden sie in Experimenten mit dem Intrusionsparadigma.

Im Intrusionsparadigma wird zwischen der Darbietung eines Satzes und seiner Wiedergabe eine Distraktoraufgabe gestellt. In dieser Aufgabe wird dem Versuchsteilnehmer eine Liste von Wörtern präsentiert und abgefragt, ob ein danach präsentiertes Wort Teil der Liste war. Daneben enthält die Liste ein Wort (*Lure*-Wort), das einem Wort im Satz eng verwandt ist und besser in den Kontext des Satzes passt, als das eigentliche Wort, das im Satz präsentiert wurde. Das Beispiel 1 (aus Rummer, 2003) soll diesen Ablauf verdeutlichen.

Satz:	Der Bericht wurde von vier Leuten geprüft, bevor der Irrtum bemerkt wurde.
Liste mit Lure-Wort:	Wolke Fehler Puder Scheune Faser
Liste ohne Lure-Wort:	Wolke Auto Puder Scheune Faser
Wortentscheidung:	Puder / Fohlen
Satzwiedergabe:	Der Bericht wurde von vier Leuten geprüft, bevor der Fehler bemerkt wurde.

Beispiel 1: Ablauf des Intrusionsparadigmas (aus Rummer, 2003)

Es zeigt sich, dass die Häufigkeit der Intrusionen des *Lure*-Wortes in den Satz ansteigt, wenn es Teil der zuvor präsentierten Liste war, also eine Voraktivierung erfahren hat. Das ursprüngliche Wort und das *Lure*-Wort sind beide durch die vorherige Präsentation voraktiviert und während der Regeneration des Satzes wird nun häufig das *Lure*-Wort selektiert. Dies ist ein Hinweis darauf, dass nicht eine Wortkette im Arbeitsgedächtnis gehalten wird, sondern die darin enthaltene Botschaft, welche bei der Wiedergabe anhand der Sprachproduktionsprozesse in einem Satz ausgedrückt wird. Dabei sind der propositionale Gehalt des Satzes und die Voraktivierung der lexikalischen Einheiten entscheidend für die Wiedergabe.

Entgegen der ursprünglichen Regenerierungshypothese finden Potter und Lombardi (1998) Hinweise auf den Einfluss der syntaktischen Struktur auf die Satzwiedergabe: Auch wenn mehrere syntaktische Strukturen zur Verfügung stehen, um einen Inhalt auszudrücken, wird die jeweils zuvor rezipierte Struktur bevorzugt. Potter und Lombardi führen dies auf syntaktisches Priming zurück, also die Voraktivierung einer syntaktischen Struktur, die somit analog zu den lexikalischen Einheiten eine höhere Aktivierung hält und damit mit größerer Wahrscheinlichkeit im Produktionsprozess selektiert wird. Entgegen Potter und Lombardis ursprünglicher Annahme finden Rummer und Engelkamp (2001, 2003, 2011; s. auch Rummer, 2003) jedoch einen Einfluss der phonologischen Information: Die Autoren erweiterten das oben beschriebene Intrusionsparadigma von Potter und Lombardi (1990) um eine Variation der Darbietungsmodalitäten (*rapid visual serial presentation* vs. auditive Präsentation) und der Wiedergabearten (direkte Wiedergabe vs. zeitliche verzögerte Wiedergabe). Dabei fanden sie keine Intrusionen in der auditiven Präsentationsbedingung, wenn die Sätze direkt wiedergegeben werden sollten, jedoch in allen anderen Bedingungen. Rummer und Engelkamp (2011) führten dies darauf zurück, dass nur bei auditiver Präsentation und direkter Wiedergabe noch phonologische Information aus dem Originalsatz verfügbar ist, nicht aber in den anderen Bedingungen. Sie argumentieren, dass phonologische Information bei der Satzwiedergabe genutzt wird, sofern sie verfügbar ist.

Schweppe (2006) fand, ebenfalls mit dem Intrusionsparadigma, einen Einfluss des grammatischen Geschlechts auf die Satzwiedergabe: Wurden den Probanden *Lure*-Wörter präsentiert, die das gleiche grammatische Geschlecht hatten wie das Zielwort, so wurden diese häufiger in der Satzwiedergabe verwendet als präsentierte *Lure*-Wörter, die ein anderes grammatisches Geschlecht hatten. Ein Einfluss durch Repräsentationen auf anderen Ebenen als der lexikalischen und konzeptuellen ist demnach wahrscheinlich. Die Oberflächeninformation könnte durchaus das Satzbehalten beeinflussen, wahrscheinlich indem ihre Aufrechterhaltung einen Teil der verfügbaren Aufmerksamkeit beansprucht. Rummer und Schweppe (2009) haben das Modell von Potter und Lombardi entsprechend erweitert.

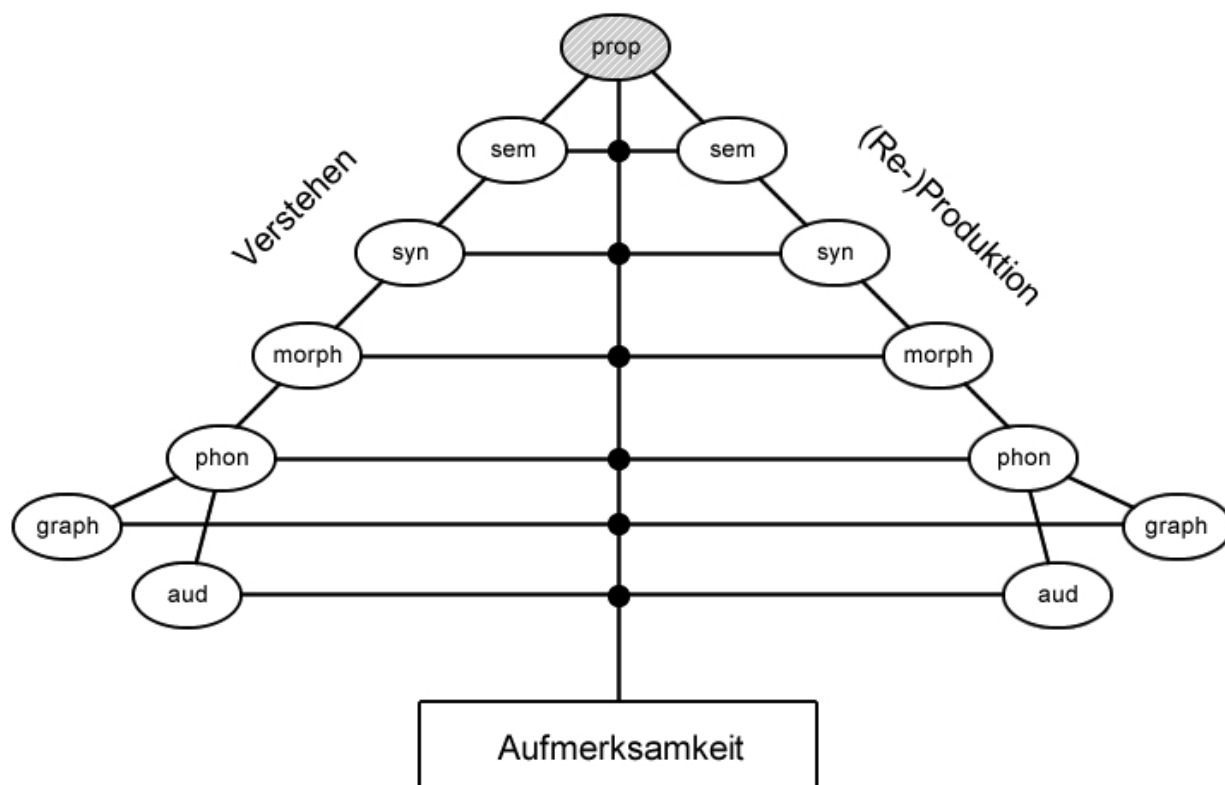


Abbildung 4: Schematische Darstellung des Rahmenmodells zum Satzbehalten von Rummer und Schweppe (2009).

In ihrem Rahmenmodell des Satzbehaltens kombinieren Rummer und Schweppe (2009) die Konzeption des Arbeitsgedächtnisses als Aktivierungszustand in einem Netzwerk und der Regeneration von Sätzen aus ihrem propositionalen Gehalt bei der Satz wiedergabe (siehe auch Schweppe & Rummer, 2007; Schweppe, Rummer & Fürstenberg, 2009; Rummer, 2003; Schweppe 2006, Rummer & Engelkamp, 2011). Sie verbinden also die oben ausgeführten Annahmen von Potter und Lombardi (1990) mit Cowans (1999) Modell des Arbeitsgedächtnisses und ergänzen diese Konzepte mit ihrer Annahme, dass neben der Informationen auf der konzeptuellen und lexikalischen Ebene der Satzverarbeitung auch die Restaktivierung auf den anderen Ebenen einen Einfluss auf die Wiedergabe des Satzes hat. Beim Verstehen eines Satzes (ohne Behaltensaufgabe) wird zunächst die eingehende Oberflächeninformation sukzessive auf der phonologischen, morpho-syntaktischen, syntaktischen und lexikalisch-semanticen Ebene analysiert. Aus diesen Informationen wird der propositionale Gehalt des Satzes extrahiert. Dieser ist auf der konzeptuellen Ebene repräsentiert. Für die Bewältigung der Mehrzahl kommunikativer Aufgaben ist der propositionale Gehalt des Satzes hinreichend, zum Beispiel für die Generierung einer Antwort, die Ausführung einer Handlung oder das Schlussfolgern auf Basis des propositionalen Gehalts. Die Oberflächenstruktur wird daher im Standardfall nicht aktiv gehalten und zerfällt schnell nach der Interpretation des Satzes. Die Produktion

eines Satzes erfolgt aus der konzeptuellen Botschaft, die syntaktisch, lexikalisch und zuletzt phonologisch realisiert wird. Dabei wird gemäß Garrett (1984) die konzeptuelle Botschaft zunächst in ihre funktionalen Rollen im Satz zerlegt, welchen dann im zweiten Schritt lexikalische Eigenschaften und eine entsprechende syntaktische Struktur (in Abhängigkeit voneinander) zugewiesen werden. Die Wörter werden dann phonologisch kodiert und entsprechend artikulatorisch umgesetzt (vgl. Vigliocco & Hartsuiker, 2002; Hartsuiker, Kolk & Huiskamp, 1999). Diese Verarbeitungsschritte sind – in der Muttersprache – weitgehend automatisiert, sie beanspruchen also nur einen sehr geringen bis gar keinen Teil der Aufmerksamkeit.

Die Verarbeitung eines Satzes zur wörtlichen Wiedergabe unterscheidet sich von diesem Prozess. Sowohl während des Verstehens als auch während der späteren Reproduktion des Satzes muss die Oberflächenstruktur aktiv gehalten werden. Dabei wird die Ressource Aufmerksamkeit bis an ihre Grenze (und je nach Satzlänge darüber hinaus) gefordert, was unter Umständen zu einem Verlust an Informationen und damit zur fehlerhaften oder unvollständigen Reproduktion führt. Die höhere Belastung besteht bereits während des Lesens eines Satzes, wie die Ergebnisse von Aaronson und Scarborough (1976) zeigen: Probanden, die angewiesen wurden, einen Satz zu lesen und danach wörtlich wiederzugeben, lasen langsamer als Probanden, die den Satz lediglich lesen sollten, um ihn zu verstehen. Daneben verlangsamte sich die Lesegeschwindigkeit in der Gruppe, die die Behaltensaufgabe erhalten hatte, über den Satz hinweg, während sie sich in der anderen Gruppe (Verständnisaufgabe) beschleunigte. Aaronson und Scarborough (1976) führten diese Befunde auf die steigende Belastung der Gedächtnisressource zurück: Je mehr Inhalt aufwändig gespeichert werden muss, desto weniger Aktivierung steht für die Verarbeitung und das Behalten zusätzlichen sprachlichen Materials zur Verfügung. Die Verarbeitung des sprachlichen Materials erfolgt daher weniger flüssig als ohne die Gedächtnisbelastung, was zu längeren Lesezeiten führt. Auch während der Reproduktion des Satzes bleibt die Aufmerksamkeit stark beansprucht, da bis unmittelbar vor ihrem Ende Oberflächeninformation aktiv gehalten werden muss. Der Satz wird aus dem propositionalen Gehalt regeneriert, dabei haben diejenigen Teile der Oberflächenstruktur die besten Chancen produziert zu werden, die aktiv gehalten wurden.

Während Potter und Lombardi annehmen, dass bei der Regeneration von Sätzen die Voraktivierung der lexikalischen Einheiten entscheidend ist, sind es bei Rummer und Schuppe Repräsentationen auf den verschiedenen Ebenen der Sprachverarbeitung, die

aktiv gehalten werden müssen. Dies ist der zentrale Unterschied zwischen den beiden Modellannahmen. Hinweise für ihre Annahme finden Schweppe (2006; s. auch Schweppe & Rummer, 2007) für morpho-syntaktische Repräsentationen und Rummer und Engelkamp (2001, 2003), sowie Park & Martin (2002) für phonologische Repräsentationen. Die Befunde von Potter und Lombardi (1998), die von diesen als Effekte syntaktischen Primings interpretiert werden, sind Hinweis auf die Aufrechterhaltung von Aktivierung auf der Ebene der syntaktischen Repräsentationen. Eine solche Beteiligung wird von Rummer, Engelkamp und Konieczny (2003) belegt.

Die Aktivhaltung der Satzoberflächeninformation ist im Modell von Rummer und Schweppe ein kontrollierter Prozess, benötigt also Aufmerksamkeit. Die begrenzte Ressource Aufmerksamkeit verteilt die Aktivierungsenergie beim Satzbehalten bereits während der Rezeption des Satzes (denn schon dann beginnt die Behaltensleistung für die ersten Bestandteile des Satzes) und auch während der Reproduktion auf viele Repräsentationen auf den Ebenen der Sprachverarbeitung.

Die Unterscheidung zwischen Verarbeitungsprozessen mit dem Ziel des Verstehens und des Behaltens findet sich auch in den Überlegungen anderer Autoren wieder. In vielen Modellen zur Verarbeitung von Sprache (z.B. Jescheniak & Levelt, 1994, Roelofs, 1992; s. jedoch Caramazza, 1997) wird der reine Verstehensprozess von Wörtern und Sätzen als hochautomatisierter Vorgang angenommen. Muss jedoch während dieses Prozesses auch Information gespeichert werden, so kann dies den Prozess beeinflussen (vgl. Just & Carpenter, 1992; s. jedoch Caplan & Waters, 1999). Dieses Zusammenspiel von automatisierter Sprachverarbeitung (die für das Verstehen von Sätzen zumeist hinreichend ist) und kapazitativ begrenzten kontrollierten Prozessen, die die Aktivierungsenergie steuern, findet sich in vielen Annahmen zum Satzbehalten wieder.

Im Folgenden werden beispielhaft die Annahmen und Studien von Kintsch und van Dijk (1978) zum Verstehen von Texten und von Just und Carpenter (1992) und Caplan und Waters (1999) zur Satzverarbeitung angeführt. Diese sollen die Rolle der Aufmerksamkeit, wie sie von Rummer und Schweppe (2005) postuliert wird und für die Annahmen in dieser Arbeit zentral ist, weiter untermauern.

Kintsch und van Dijk (1978) beschreiben komplexe Prozesse des Verstehens und der Produktion von Texten auf der Ebene der semantischen Struktur. Sie gehen von einer weitgehend automatisierten Verarbeitung beim Verstehen von gehörten und gelesenen Texten aus: „...we assume several complex processes operating in parallel and

interactively without straining the resources of the system" (Kintsch & van Dijk, 1978, p. 364). Auf der semantischen Ebene kann ein Text in Propositionen zerlegt werden, die im Text miteinander verbunden sind und Bezug aufeinander nehmen. Diese Bezüge bestehen nach Kintsch und van Dijk (1978) auf den beiden Ebenen der Mikrostrukturen und der Makrostrukturen. Auf der Ebene der Mikrostrukturen werden die Propositionen entsprechend ihrer Koreferenzen und ihrer Hierarchie geordnet. Die Makroebene verbindet diese Propositionen gemäß des Themas des Diskurses, also über die lokalen Verbindungen hinaus. Dieses Modell der Textverarbeitung enthält Aussagen nur über Teilprozesse der Satzverarbeitung. Wichtig für diese Arbeit ist ihre Annahme, dass der Prozess des Verstehens selbst nicht belastend für die Ressourcen des Systems ist, die Speicherung von Information jedoch sehr wohl, wie Kintsch und van Dijk (1978, S. 364) betonen: "Capacity limitations become crucial, however, when it comes to the storage of information in memory and response production".

In ihrem Modell der *Capacity Constrained Comprehension* beschreiben Just und Carpenter (1992) ein Modell des verbalen Arbeitsgedächtnisses, das sich aus verschiedenen Prozessen der Sprachverarbeitung zusammensetzt. Für diese Prozesse steht eine begrenzte Menge an Aktivierungsenergie zur Verfügung. Ein Überschreiten der limitierten Kapazitäten des verbalen Arbeitsgedächtnisses entweder durch Verarbeitungs-, Speicher- oder Manipulationsprozesse führt zu einer Verlangsamung der Verarbeitung und/oder einem Verlust verarbeiteter Information: „The scheme implies that when the task demands are high (either because of storage or computational needs), then processing will slow down and some partial results may be forgotten. [...] When the task demands exceed the available resources, both storage and computational functions are degraded“ (Just & Carpenter, 1992, S. 123-124). In Just und Carpenters Modell greifen also Prozesse des reinen Sprachverstehens und Operationen mit den sprachlichen Einheiten auf die selbe Ressource zurück, so dass es durch erhöhte Anforderungen in den Operationen zu Interferenz mit der Sprachverarbeitung (und umgekehrt) kommen kann. Muss ein Satz nicht nur verstanden, sondern auch behalten werden, wird die Aktivierungsenergie gleichzeitig für Sprachverarbeitungs- und Speicherprozesse benötigt. Die Verarbeitung eines Satzes zum Behalten benötigt also mehr Aktivierungsenergie, da gleichzeitig der Verstehens- und der Speicherprozess Energie benötigen. Je höher die Anforderungen an die Behaltensleistung werden, desto weniger Energie steht für die Verarbeitung zur Verfügung. Umgekehrt kann die Komplexität des zu verarbeitenden Materials die Behaltensleistung beeinflussen. Je sprachlich komplexer ein Satz ist, desto mehr

Behaltensleistung muss erbracht werden, um die einzelnen Bestandteile des Satzes zu einer konzeptuellen Botschaft zu integrieren.

Caplan und Waters (1999) gehen im Gegensatz dazu von einer geteilten Ressource aus. In ihrer *Separate-Sentence-Interpretation-Resource-Hypothese* (SSRI-Hypothese) formulieren sie die Annahme, dass die Prozesse der Satzinterpretation, die sie als automatisiert betrachten, und die kontrollierte Verarbeitung verbalen Materials (z.B. Behaltensleistungen, Schlussfolgern) auf unterschiedliche Ressourcen zugreifen. Sie unterscheiden Prozesse danach, ob sie interpretativ (*interpretive processing*) oder postinterpretativ (*post-interpretive processing*) sind. Zu den interpretativen Prozessen gehören linguistische Prozesse, die die reine Verarbeitung des Satzes betreffen, wie das Erkennen von Wörtern, und die Zuweisung ihrer Bedeutung und syntaktischen Eigenschaften, sowie die semantische Interpretation des Satzes. Postinterpretative Prozesse sind solche, in denen die in den interpretativen Prozessen verarbeitete Information angewandt wird. Das kann das Behalten des Satzes sein, aber auch die Planung einer Aktion basierend auf der Bedeutung des Satzes oder das Nachdenken über den Inhalt (vgl. Caplan & Waters, 1999). Caplan und Waters (1999) argumentieren, dass für die interpretativen Prozesse eine spezielle Ressource des Arbeitsgedächtnisses zur Verfügung steht. Diese ist zwar Teil des verbalen Arbeitsgedächtnisses, jedoch in einem Subsystem von der Ressource getrennt, die für die postinterpretativen Prozesse zur Verfügung steht. Auf Grundlage des Modells von Caplan und Waters (1999) können Unterschiede zwischen Sprechern einer L1 und L2 im Satzbehalten scheinbar nicht durch ihre in unterschiedlichem Maße automatisierte Sprachverarbeitung erklärt werden, da die Ressourcen für die Sprachverarbeitung und das Behalten eines Satzes separat voneinander sind, eine stärkere Belastung der einen Ressource also nicht zu einer geringeren Leistung in der anderen führen sollte. Jedoch könnte das Lesen eines Satzes, mit dem Ziel, ihn wörtlich wiederzugeben, einen Sonderfall der Verarbeitung darstellen. Caplan and Waters (1999, S. 93) betonen: "Listeners appear to accomplish sentence interpretation in two ways, in the usual obligatory, unconscious and fast mode, or in a mode in which conscious, controlled processes are applied to the task. We do not conceive of the latter type of processing as belonging to the set of operations that we suggest utilize a specialized resource pool".

Ein experimenteller Hinweis auf eben diesen Verarbeitungsunterschied zwischen dem Lesen eines Satzes, der wörtlich wiedergegeben werden sollte und dem Lesen eines

Satzes, der lediglich verstanden werden sollte fanden Aaronson und Scarborough (1976) in ihrer bereits mehrfach erwähnten Lesestudie. Ihre Befunde sind im Einklang mit Just und Carpenters (1992) Modell der *Capacity Constrained Comprehension*. Sie können auch mit den Annahmen von Caplan und Waters (1999) vereinbart werden, wenn das Verarbeiten eines Satzes zur darauf folgenden Wiedergabe als ein Fall der kontrollierten Verarbeitung interpretiert wird, der nicht in die Teilmenge der interpretativen Prozesse fällt. Der Sonderfall der Verarbeitung eines Satzes zur Wiedergabe ist in Caplan und Waters (1999) Modell vermutlich so zu modellieren, dass interpretative Prozesse *und* postinterpretative Prozesse an der Verarbeitung beteiligt sind. Problematischer ist es, eine Vorhersage für den Vergleich der Leistung von Muttersprachlern und *Near-Natives* in ihrer Leistung im Satzbehalten innerhalb des Modells von Caplan und Waters (1999) zu machen. Wenn die oben berichteten Unterschiede zwischen Sprechern einer L1 und einer L2 auf dieses Modell übertragen werden, so ist anzunehmen, dass sich Sprecher einer L1 und einer L2 im Automatisierungsgrad der interpretativen Prozesse unterscheiden, aber nicht in den postinterpretativen. Da jedoch durch das Satzbehalten nur die postinterpretativen Prozesse in besonderer Weise gefordert sind (durch die Behaltensanforderung) sollten Unterschiede im Automatisierungsgrad der interpretativen Prozesse nicht zu einem Leistungsunterschied in der Wiedergabe führen: Aufgrund der Trennung der beiden Ressourcen kann ein geringerer Automatisierungsgrad der interpretativen Prozesse nicht durch die Ressource der postinterpretativen Prozesse ausgeglichen werden. Ob und wie genau sich die geringere Automatisierung der interpretativen Prozesse auf die Leistung in der Satzwiedergabe auswirken würde, kann aus den Modellannahmen von Caplan und Waters (1999) nicht direkt abgeleitet werden.

Laut Just und Carpenter (1992) wäre es die geteilte Ressource des verbalen Arbeitsgedächtnisses, deren Kapazität beim Satzbehalten (von entsprechend langen Sätzen) überschritten wird. Die Beschleunigung der Lesedauer in der Bedingung „Verständnis“ erklären Aaronson und Scarborough (1976) mit der schnelleren Verarbeitung redundanter kontextueller Informationen und der oberflächlicheren Verarbeitung der syntaktischen Struktur. Sie weisen auch darauf hin, dass Lesestrategien, die mit dem Ziel verwendet werden, einen Satz zu verstehen, sich fundamental von den Strategien zum Behalten eines Satzes unterscheiden könnten und sogar inkompatibel sein könnten: "[...] optimal coding for comprehension and verbatim memory may be incompatible in many ways" (Aaronson & Scarborough, 1976, S. 65). Wichtig ist, dass bereits die Leseprozesse je nach Schwerpunkt der Aufgaben unterschiedlich ablaufen,

obwohl sich die Aufgaben erst im zweiten Schritt (Wiedergabe bzw. Antwort auf Verständnisfrage) unterscheiden. Dieser frühe Unterschied deutet auf einen Mechanismus hin, der auf den Leseprozess bzw. die Enkodierung des Satzes Einfluss nimmt und den Einsatz der entsprechenden Strategie je nach Aufgabenstellung steuert. Die Vermutung liegt nahe, dass die Satzwiedergabe nicht als Satzverstehen plus Satzproduktion plus Arbeitsgedächtnis gedeutet werden kann, wie es bei der Verwendung in Sprachtests für Kinder geschieht. Die Interaktion der zu Grunde liegenden Prozesse ist weitaus komplexer als diese Interpretation glauben macht.

Zusammenfassend will ich noch einmal die Bedeutung betonen, die den kontrollierten Prozessen der Aufmerksamkeit im Zusammenspiel mit der eher automatisierten Sprachverarbeitung beim Satzbehalten zukommt. Alle vorgestellten Modelle des Arbeitsgedächtnisses gehen von einer begrenzten steuernden Ressource aus, die die Aufrechterhaltung von Information im Arbeitsgedächtnis beeinflusst. Sprachverarbeitung wird zumeist als stufenweiser automatisierter Prozess modelliert. Jedoch beeinflusst die Behaltensanforderung in allen Modellen den Verstehensprozess maßgeblich, was Aaronson und Scarborough (1976) auch experimentell zeigen konnten. Zuletzt soll festgehalten werden, dass entgegen Potter und Lombardis (1990, s.a. Lombardi & Potter, 1992) ursprünglicher Annahme ein Einfluss der Oberflächeninformation bei der Satzwiedergabe aufgezeigt werden kann. Oberflächeninformation beeinflusst demnach wahrscheinlich den Wiedergabeprozess beim Satzbehalten.

5 Satzbehalten in einer Zweitsprache - Forschungsfrage und Hypothesen

In diesem Kapitel sollen die zuvor ausführlich referierten Modellannahmen zusammengeführt und noch einmal explizit auf die Fragestellung dieser Arbeit bezogen werden. Es wird diskutiert, welche Prozesse am Satzbehalten in einer L1 und einer L2 beteiligt sind und welche Probleme dies für den Vergleich der Leistung der jeweiligen Sprecher in dieser Aufgabe mit sich bringt. Dann werden daraus die zentralen Annahmen dieser Arbeit abgeleitet und abschließend die im nächsten Kapitel berichteten Studien skizziert.

Die in dieser Arbeit vertretenen zentralen Annahmen lauten folgendermaßen: Die Verarbeitung einer L1 und L2 unterscheidet sich selbst bei hochkompetenten L2-Sprechern im Grad ihrer Automatisierung. Bei der Satz wiedergabe in einer L1 können daher die Ressourcen der Aufmerksamkeit in höherem Maße auf die Aktivhaltung der Satzoberfläche konzentriert werden, als dies für eine L2 möglich ist. Folgt man Just und Carpenter (1992) in der Annahme einer universellen Aufmerksamkeitsressource, die für Prozesse im verbalen Arbeitsgedächtnis zur Verfügung steht, so teilt sich diese beim Satzbehalten auf die Verarbeitung des Satzes und die Aktivhaltung der Satzoberflächeninformation und des Satzinhalts auf. Je weniger Aufmerksamkeit für die Satzverarbeitung benötigt wird, desto mehr Aufmerksamkeit kann auf die Aktivierung der Oberflächeninformation verwendet werden und desto besser sollte diese in der Wiedergabe erhalten bleiben. Die Automatisierung der Sprachverarbeitung erfolgt zumindest für einige Ebenen (insbesondere die Phonologie) bereits im frühen Kindesalter (vgl. Sebastián-Gallés & Bosch, 2005; Sebastián-Gallés, Echeverría & Bosch, 2005). Bei einem später einsetzenden L2-Erwerb kann eine muttersprachliche Verarbeitung auch durch Sprachförderung nicht mehr erreicht werden. Abseits der muttersprachlichen Verarbeitung kann jedoch eine muttersprachliche Kompetenz in der L2 auch noch mit einem mittleren L2-Erwerbsalter (AoA < 16) erreicht werden. Diese Kompetenz kann erreicht werden, obwohl die Verarbeitungsprozesse in ihrem Automatisierungsgrad von denen der Muttersprachler abweichen. Dabei ist der Automatisierungsgrad einer Sprache generell als Teilkomponente der Sprachkompetenz zu verstehen: Mit einer niedrigen Sprachkompetenz geht immer auch ein niedriger Grad an Automatisierung einher. Dennoch sind ab dem sehr hohen Level von Automatisierung, das *Near-Natives* erreicht haben, Unterschiede im Automatisierungsgrad im Vergleich zu Muttersprachlern nicht mehr sinnvoll als Kompetenzunterschiede zu interpretieren. Zeigt ein L2-Sprecher muttersprachliche

Sprachkompetenz, so soll dieses gezeigte Verhalten ausschlaggebend für die Beurteilung der Kompetenz sein, nicht der von den muttersprachlichen Prozessen leicht abweichende Automatisierungsgrad, mit der diese Kompetenz hervorgebracht wird.

Die Satzwiedergabeleistung als Maß der Sprachkompetenz betont diese Automatisierungsunterschiede jedoch überproportional. Beim Satzbehalten wird bereits während des Verstehensprozesses die Botschaft des Satzes und seine Oberflächenstruktur als Information aktiv gehalten. Da die Sätze bei der Aufgabe "Sätze wiedergeben" so lang sind, dass auch Muttersprachler sie meist nicht vollständig wiedergeben können, wird die Aufmerksamkeitsressource bis über ihre Grenzen hinaus beansprucht. Dies zeigt sich im Informationsverlust zwischen Rezeption und Reproduktion. Kann weniger Kapazität für die Behaltensaufgabe bereitgestellt werden, weil mehr Kapazität für die Verarbeitung des Satzes benötigt wird, so erhöht sich der zu erwartende Verlust an Information. Es werden weniger Wörter des Satzes korrekt wiedergegeben, die Leistung in der Aufgabe sinkt. Dies ist der Fall beim Testen in einer L2, da hier mehr Aufmerksamkeit auf die Verarbeitung des Satzes verwendet werden muss als in einer L1. Die Ursache hierfür ist der geringere Grad der Automatisierung einer L2 im Vergleich zu einer L1. Damit steht in einer L2 weniger Kapazität für das Behalten des Satzes zur Verfügung als in einer L1, was zu schlechterer Leistung in der Satzwiedergabe für *Near-Natives* im Vergleich zu Muttersprachlern führt.

Ich möchte noch einmal betonen, dass diese schlechtere Leistung nicht als Kompetenzunterschied gedeutet werden sollte, sondern vielmehr als reiner Verarbeitungsunterschied. In den Kapiteln 2 und 3 wurde dargelegt, dass nahezu muttersprachliche Kompetenz nicht in jedem Fall mit muttersprachlichen Verarbeitungsprozessen einhergeht. Bei Sprechern einer L2 mit sukzessivem Spracherwerb konnten durch neuropsychologische Methoden Automatisierungsunterschiede in Abwesenheit von Kompetenzunterschieden aufgezeigt werden. Verarbeitungsunterschiede auf der phonologischen Ebene zwischen den beiden Sprachen von simultan bilingual aufgewachsenen Erwachsenen konnten experimentell durch Vokalkontraste aufgezeigt werden. Obwohl diese Unterschiede aufzeigbar waren, sind die simultan bilingual aufgewachsenen Sprecher Muttersprachler in beiden Sprachen und bewegen sich als solche in der Sprachgemeinschaft. Dies demonstriert, dass Verarbeitungsunterschiede nicht in jedem Fall als Kompetenzunterschiede zu deuten sind.

Die in den nächsten Kapiteln berichteten Studien lassen sich zwei verschiedenen Zielen zuordnen: Im ersten Teil sollen die grundlegenden Annahmen dieser Arbeit geprüft werden. Im zweiten Teil soll untersucht werden, ob eine Modifikation der Satzwiedergabe, die vermutlich geringere Anforderungen an das verbale Arbeitsgedächtnis stellt, dazu führt, dass die Verarbeitungsunterschiede zugunsten der wahren Kompetenzunterschiede in den Hintergrund treten. Im Folgenden wird Logik und Struktur dieser beiden Teile detailliert vorgestellt.

Im ersten Teil dieser Arbeit soll getestet werden, ob die Leistung in der Satzwiedergabe tatsächlich durch Verarbeitungsunterschiede beeinflusst wird, die nicht als Kompetenzunterschiede zu werten sind. Diese Verarbeitungsunterschiede führen dazu, dass hochkompetente Sprecher einer L2 schlechter in der Satzwiedergabe abschneiden, als aus ihrer Leistung in Vergleichsmaßen der Sprachkompetenz und der Leistung von Muttersprachlern in beiden Maßen (Satzwiedergabe und Vergleichsmaß) zu erwarten ist. Dies soll in drei quasi-experimentellen Studien getestet werden. Alle Studien vergleichen die Leistung hochkompetenter Sprecher einer L2 in der Satzwiedergabe und einem Vergleichsmaß der Sprachkompetenz mit der Leistung von Muttersprachlern in diesen beiden Maßen. Um zu zeigen, dass dieser Effekt für die Zielgruppe der Sprachstandserhebungen, also vierjährige Kinder, zutrifft, vergleicht die erste Studie die Leistung von Kindergartenkindern mit DaM und DaZ im Nachsprechen von Sätzen und in anderen sprachbezogenen Aufgaben. Diese Aufgaben sind dem SETK 3-5 (Grimm, 2001), einem gängigen Sprachentwicklungstest für die Altersstufe der 3-5-Jährigen, entnommen. Dabei wird erwartet, dass sich kompetente Kinder mit DaZ in ihrer Leistung im Nachsprechen von Sätzen stärker von gleichaltrigen Kindern mit DaM unterscheiden als im Vergleichsmaß. Dies wäre ein erster experimenteller Hinweis darauf, dass durch die Satzwiedergabe die Leistung von Kindern mit DaZ unterschätzt wird. Da diese Kinder aber den Prozess des L2-Erwerbs (und auch des L1-Erwerbs) noch nicht abgeschlossen haben, bestehen allgemeine Kompetenzunterschiede zwischen den Kindern mit DaZ und DaM. Ziel dieser Arbeit ist jedoch, Unterschiede in der Leistung in der Satzwiedergabe in Abwesenheit von Kompetenzunterschieden aufzuzeigen.

Dazu werden in der zweiten Studie Muttersprachler und *Near-Natives* des Deutschen in ihrer Leistung in der Wiedergabe von Sätzen und in einem Vergleichsmaß der allgemeinen Sprachkompetenz, dem C-Test, verglichen. Der C-Test ist ein Lückentext, der wie die Satzwiedergabe die Rezeption, Produktion und (mit Einschränkungen) auch die Automatisierung der Sprachverarbeitung testet. Er ist eines der wenigen Testverfahren, das

mit Muttersprachlern und Zweitsprachlern eingesetzt werden kann. Darüber hinaus verlangt er ähnliche Fähigkeiten wie die Satzwiedergabe, belastet das Arbeitsgedächtnis jedoch nicht mit einer Behaltensanforderung. Es wird erwartet, dass *Near-Natives*, die die gleiche Leistung wie Muttersprachler im C-Test aufweisen, schlechter als Muttersprachler in der wörtlichen Satzwiedergabe abschneiden.

Da bis zu diesem Punkt ausschließlich monolinguale Probanden mit bilingualen verglichen werden, soll in Studie 3 gezeigt werden, dass die geringere Automatisierung auf die L2 beschränkt ist und dass es sich dabei nicht im einen Nachteil für bilinguale Sprecher per se handelt, der sich in beiden Sprachen niederschlägt. Außerdem soll die Übertragbarkeit des in den Studien 1 und 2 aufgezeigten Effekts auf eine andere Sprache, das Französische, demonstriert werden. Dazu wird eine Gruppe von Schülern, die teilweise Deutsch-Französisch, teilweise Französisch-Deutsch bilingual sind, in ihren beiden Sprachen anhand der Satzwiedergabe und dem C-Test als Vergleichsmaß untersucht. Analog zur ersten Studie wird erwartet, dass sich die Schüler in der Satzwiedergabe in ihrer jeweiligen L2 stärker von Sprechern der jeweiligen L1 unterscheiden als im Vergleichsmaß der Sprachkompetenz.

Im zweiten Teil der Arbeit soll versucht werden, eine alternative Testaufgabe zum fairen Testen von Muttersprachlern und Zweitsprachlern zu identifizieren. Diese Testaufgabe sollte die Vorzüge der Satzwiedergabe beibehalten, aber die Behaltensanforderung reduzieren. Dies wäre der Fall, wenn die Satzwiedergabe als Aufgabentyp grundsätzlich beibehalten würde, aber die Behaltensanforderung sich auf den Inhalt des Satzes beschränken würde (statt Inhalt *und* Satzoberfläche). In Studie 4 wird getestet, ob diese Modifikation der Aufgabenstellung tatsächlich zu einer niedrigeren Belastung des Arbeitsgedächtnisses führt. Dazu wird das Paradigma der bereits beschriebenen Lesestudie von Aaronson und Scarborough (1976) adaptiert. Diesem Paradigma wird in Studie 4 neben dem Lesen zum Verständnis und der wörtlichen Satzwiedergabe eine weitere Bedingung hinzugefügt, die der inhaltlichen Satzwiedergabe. Wie in Aaronson und Scarboroughs (1976) Studie wird die Lesegeschwindigkeit dabei als Indikator für die Belastung des Arbeitsgedächtnisses interpretiert. Es wird erwartet, dass die Lesegeschwindigkeit mit inhaltlicher Behaltensanforderung höher ist als mit wörtlicher Behaltensanforderung, jedoch geringer als beim Lesen ohne Behaltensanforderung.

Konnte die geringere Belastung beim inhaltlichen Behalten im Vergleich zum wörtlichen Behalten belegt werden, so soll abschließend untersucht werden, ob die aufgabenspezifischen Leistungsunterschiede zwischen Muttersprachlern und *Near-Natives* verschwinden, wenn die Satzwiedergabe mit einer geringeren Belastung einhergeht, wenn also die Sätze nur inhaltlich und nicht wörtlich behalten werden müssen. Dies wird in Studie 5 mit erwachsenen *Near-Natives* getestet, deren Leistung in der inhaltlichen bzw. wörtlichen Satzwiedergabe und dem C-Test als Vergleichsmaß der Sprachkompetenz wiederum mit der Leistung von Muttersprachlern in diesen Maßen verglichen wird.

6 Studie 1 - Sätze nachsprechen bei Vierjährigen

In dieser ersten Studie soll untersucht werden, ob zwischen Kindern mit DaZ, die ihre L2 Deutsch gut beherrschen, und Kindern mit DaM in der Aufgabe "Sätze nachsprechen" ein Unterschied gefunden werden kann, der über den erwarteten Nachteil durch ihre kürzere L2-Spracherwerbsdauer hinausgeht. Genauer gesagt soll überprüft werden, ob die Kinder mit DaZ im Nachsprechen von Sätzen schlechter abschneiden, als sich aus ihrer Leistung in alternativen sprachlichen Aufgabenstellungen und der Leistung von Kindern mit DaM in diesen Aufgabenstellungen und dem Nachsprechen von Sätzen ableiten lässt.

Es wird erwartet, dass Kinder mit DaZ in den Aufgaben insgesamt hinter der Leistung der muttersprachlichen Kinder zurückbleiben, da sie zum Testzeitpunkt weniger Kontaktmonate mit der deutschen Sprache hatten als die Gruppe der muttersprachlichen Kinder. Dabei sollte sich die Leistung der beiden Gruppen im Aufgabentyp "Sätze nachsprechen" deutlicher unterscheiden als in den übrigen Untertests des SETK 3-5, die wie oben bereits erwähnt als eher geeignet für den Einsatz mit Kindern mit DaZ gelten (vgl. Kany & Schöler, 2010). Der Aufgabentyp "Sätze nachsprechen" wird dabei in zwei Versionen getestet: Einmal in der Version des SETK 3-5 (zehn Sätze von jeweils sechs bis neun Wörtern Länge) und zum anderen in der Version von Delfin 4 (zwei Sätze, sechs und neun Wörter Länge)⁹. Dabei wird erwartet, dass die Leistungsunterschiede zwischen den Gruppen in den beiden Versionen der Aufgabe "Sätze nachsprechen" größer sind, als die Unterschiede zwischen den Gruppen in den übrigen Untertests. In einem 2x2-Design werden die Faktoren L1 (DaM vs. DaZ) und TEST ("Sätze nachsprechen" vs. SETK 3-5 Untertests 1, 3, 4) betrachtet¹⁰. Der erwartete deutlichere Unterschied im "Sätze nachsprechen" zwischen den Gruppen sollte sich in diesem 2x2 Design in einer Interaktion zwischen den Faktoren L1 (DaM vs. DaZ) und TEST ("Sätze nachsprechen" vs. SETK 3-5 Untertests 1, 3, 4) zeigen.

⁹ Sowohl der SETK 3-5, als auch Delfin 4 verwenden neben sinnvollen Sätzen auch sinnfreie Sätze. Diese werden, aus Gründen, die weiter oben ausführlich dargelegt wurden, in dieser Arbeit nicht betrachtet und fließen nicht in die Vergleiche mit ein.

¹⁰ Die Berechnung eines kombinierten Maßes für die Untertests 1, 3 und 4 des SETK 3-5 wird im Folgenden ausführlich beschrieben.

6.1 Methode

Material¹¹

SETK 3-5

Der SETK 3-5 (Grimm, 2001) in der hier verwendeten Version für Kinder von 4;00 bis 5;11 Jahre besteht aus 5 Untertests, von denen der fünfte ein Kontrolltest ist und nicht in die Wertung einfließt. Im Folgenden werden die Untertests detailliert beschrieben.

Der Untertest 1 „Verstehen von Sätzen“ testet das Verständnis für zunehmend komplexere syntaktische Konstruktionen anhand einer Manipulationsaufgabe. Kinder hören Sätze und sollen entsprechende Objekte, die auf dem Tisch vor ihnen liegen, gemäß des Satzinhalts bewegen (z.B. „Leg den Teddy neben das Bilderbuch.“). Für jeden Satz, der richtig umgesetzt wird, wird ein Punkt vergeben; es kann ein Höchstwert von 15 Punkten erreicht werden.

Im Untertest 2 „Satzgedächtnis“ werden den Kindern Sätze vorgelesen, die sie nachsprechen sollen. Im ersten Teil sind dies zehn sinnvolle Sätze mit je sechs bis zehn Wörtern Länge (z.B. „Die Ente sitzt neben dem Auto.“), im zweiten Teil sind die Sätze semantisch ungewöhnlich und „sinnlos“ (z.B. „Der viereckige Indianer gießt den glücklichen Kuchen in einen Sack.“). In der Aufgabe „Satzgedächtnis“ wird für jedes richtig wiederholte Wort ein Punkt vergeben. Betrachtet man nur die zehn sinnvollen Sätzen kann so ein Höchstwert von 46 Punkten erreicht werden.

Der 3. Untertest „Phonologisches Gedächtnis für Nichtwörter“ testet die Fähigkeit, mehrsilbige Kunstwörter zu wiederholen, die den phonotaktischen Regeln des Deutschen entsprechen, für jedes korrekt wiederholte Wort wird ein Punkt vergeben. Diese Kunstwörter werden zunehmend länger und komplexer (z.B. Billop, Glösterkeit, Vominlapertust). Es können maximal 18 Punkte in dieser Aufgabe erreicht werden.

Im 4. Untertest „Morphologische Regelbildung“ werden den Kindern Karten mit Bildern von Objekten präsentiert. Diese werden im Singular vom Versuchsleiter benannt, die Kinder sollen die entsprechende Pluralform benennen. Im ersten Teil der Aufgabe sind dies Objekte, die den Kindern bekannt sein sollten (z.B. Auto, Fisch), im zweiten Teil werden ihnen Kunstwörter (mit Bildern unbekannter Gegenstände) präsentiert (z.B. Tulo,

¹¹ Die Kenntnis der Materials ist Voraussetzung für das Verständnis von Aspekten der Stichprobenauswahl, daher wird entgegen der Konvention in dieser Studie die Materialbeschreibung vor die Stichprobenbeschreibung gesetzt.

Ribane). Für jeden korrekt gebildeten Plural werden 2 Punkte vergeben, 1 Punkt wird für gewisse, im Handbuch spezifizierte, Abweichungen vergeben. Insgesamt können im Untertest 4 bis zu 36 Punkte erreicht werden.

Untertest 5 „Gedächtnis für Wörter“ wird gemäß des Handbuchs des SETK 3-5 nicht in der weiteren Analyse berücksichtigt. In diesem Untertest wiederholen die Kinder eine Folge von einsilbigen, bekannten Wörtern (z.B. „Tisch Bett Stuhl“). Dabei gilt die Aufgabe als erfolgreich absolviert, wenn mindestens 3 Wörter in der korrekten Reihenfolge wiederholt werden können. Bei Kindern, die weniger als 3 Wörter wiederholen, könnte ein Defizit des phonologischen Arbeitsgedächtnisses vorliegen. Keines der Kinder, die getestet wurden, produzierte weniger als 3 Wörter in korrekter Reihenfolge.

Der SETK 3-5 ordnet den Rohwerten jeweils einen Testwert (T-Wert) zu. Diese T-Werte setzen die Rohwerte in Beziehung zu den (rein muttersprachlichen) Normen des SETK 3-5. Kinder, die als unauffällig in ihrer Entwicklung eingestuft werden, sollten einen T-Wert von mindestens 40 erreichen, der mittlere T-Wert ist für alle Untertests auf 60 genormt. Die T-Werte können nicht für die Datenanalyse herangezogen werden, da keine separaten T-Werte für die sinnvollen Sätze vorliegen. Auch ist der Vergleich mit der Aufgabe "Sätze nachsprechen" aus Delfin 4 anhand von T-Werten nicht möglich, da für Delfin 4 keine entsprechenden T-Werte vorliegen. Darüber hinaus verhindern sie den hier gewünschten Vergleich der Testschwierigkeit der einzelnen Untertests für die Gruppen DaZ und DaM durch die Normierung um den Mittelwert 60 (SD = 10).

Die im SETK 3-5 erreichten Rohwerte sind jedoch ebenso wenig für die Analyse geeignet, da die Untertests sich in ihrer maximal erreichbaren Punktzahl unterscheiden (z.B. 15 im Untertest 1 vs. 36 im Untertest 4). Um einen vergleichbaren Bezugsrahmen für die Aufgaben herzustellen, wurde daher für jeden Untertest ermittelt, wie viel Prozent der Höchstpunktzahl erreicht wurde. Für den Untertest 4 (Höchstpunktzahl 36) wäre ein Rohwert von 9 Punkten also ein Prozentwert von 25, während im Untertest 1 (Höchstwert 15) aus einem Rohwert von 3 ein Prozentwert von 20 resultiert.

Delfin 4

Wie zu Beginn dieses Kapitels bereits berichtet, wurde neben dem SETK 3-5 auch die Aufgabe „Sätze nachsprechen“ aus der Sprachstandserhebung Delfin 4 eingesetzt. Da

die Konzeption von Delfin 4 es vorsieht, jeweils vier Kinder zur gleichen Zeit zu testen, verwendet Delfin 4 vier verschiedene Satzpaare, also ein Paar für jedes getestete Kind. Der jeweils erste Satz besteht aus 6 Wörtern, der zweite aus 9 Wörtern. Bei Delfin 4 (wie auch beim SETK 3-5 Untertest "Satzgedächtnis") ergibt sich die maximal erreichbare Punktzahl im Untertest aus der Summe der Wörter. Maximal konnten also 15 Punkte (= korrekt wiedergegebene Wörter) erreicht werden. Analog zu Delfin 4 wurde jedem Kind jeweils eines der vier Satzpaare präsentiert, die Verwendung der verschiedenen Satzpaare wurde über die teilnehmenden Kinder hinweg ausbalanciert. Auch die Ergebnisse der Aufgabe "Sätze nachsprechen" aus Delfin 4 wurden für die Analyse in dieser Studie in Prozentwerte überführt.

6.1.1 Stichprobe

44 Kinder im Alter von 4;00 bis 5;03 Jahren in Kindergärten und -tagesstätten in Berlin, Bielefeld und Saarbrücken nahmen an der Studie teil. Die Daten von zwei Kindern mit DaM flossen nicht in die weitere Analyse mit ein, da bei diesen Kindern laut Auskunft der Eltern diagnostizierte Sprachstörungen vorlagen. Von den verbleibenden 42 Kindern sprachen 22 Deutsch als Muttersprache, davon 10 Mädchen, und 20 Kinder Deutsch als Zweitsprache, davon 14 Mädchen. Alle Kinder mit DaZ waren sukzessive bilingual: weder Mutter noch Vater und auch keine andere enge Bezugsperson sprachen Deutsch als Muttersprache, der Erwerb der deutschen Sprache setzte erst nach dem zweiten Lebensjahr ein. Zum Zeitpunkt der Erhebung besuchten die Kinder seit mindestens 18 Monaten regelmäßig (jeweils mehrere Stunden an 5 Tagen pro Woche, Ferienzeiten ausgenommen) eine deutsche Kindertagesstätte bzw. einen deutschen Kindergarten. Diese Informationen wurden mittels eines Fragebogens (teilweise in der L1 der Eltern) erhoben (siehe Anhang B1). Das Einverständnis der Eltern war im Vorfeld eingeholt worden, alle Kinder nahmen freiwillig an der Studie teil.

Die vorliegende Studie untersucht die Leistung sprachlich kompetenter L2-Sprecher. Daher werden nur Kinder mit DaZ in der Analyse berücksichtigt, die diese Anforderung erfüllen. Um diese Kinder zu identifizieren, wurden die T-Werte des SETK 3-5 für die Untertests 1, 3 und 4 herangezogen. Es flossen nur Daten von Kindern mit DaZ in die Analyse mit ein, die mindestens einen T-Wert von 40 (gemittelt über die drei Untertests 1, 3 und 4) erreichten. Dieser Wert ist laut der Normen des SETK 3-5 die Untergrenze für sprachlich unauffällige Kinder mit DaM. Damit flossen die Daten von 10

Kindern mit DaZ (acht Mädchen und zwei Jungen) mit einem durchschnittlichen Alter von 4,33¹² Jahren (SD = 0,28; Range 4,00 - 4,75) und von allen getesteten Kindern mit DaM mit einem durchschnittlichen Alter von 4,61 Jahren (SD = 0,33; Range 4,00 - 5,25) in die nachfolgende Analyse ein.

6.1.2 Ablauf

Die Kinder wurden jeweils mit dem SETK 3-5 (Grimm, 2001) und der Aufgabe „Sätze nachsprechen“ aus der Sprachstandserhebung Delfin 4 (Fried, 2008) getestet. Die Tests fanden in Einzelsitzungen von 20-25 Minuten Dauer in einem ruhigen Raum in den teilnehmenden Kindergärten und Kindertagesstätten statt. Die Versuchsleiterin war Muttersprachlerin des Deutschen. Die Sitzungen wurden mit einem Diktiergerät zur späteren Auswertung aufgezeichnet.

Die Delfin 4 Sätze wurden zusammen mit den SETK 3-5 Sätzen im Untertest "Satzgedächtnis" dargeboten, so dass in diesem Untertest 12 Sätze präsentiert wurden (10 aus dem SETK 3-5 und 2 aus Delfin 4). Dies ist möglich, da die Instruktionen in beiden Tests identisch sind und erscheint sinnvoll, um den Kindern eine Dopplung eines Aufgabentyps und eine doppelte Instruktion zu ersparen. In der Auswertung werden die beiden Teilaufgaben jedoch separat behandelt.

6.1.3 Auswertung

Die Sitzungen wurden mit einem Diktiergerät aufgezeichnet und später auf die SETK 3-5 Antwortbögen transkribiert. Für Untertest 1 „Sätze verstehen“ wurden die erreichten Punkte während der Sitzung direkt in die Antwortbögen eingetragen, da hier die stille Aktion, die ein Kind ausführt, beobachtet wird.

Wie bereits aus der Beschreibung des Materials ersichtlich wird, variieren die maximal erreichbaren Werte zwischen den Untertests. Daher können für die Auswertung nicht die Rohwerte berücksichtigt werden. Dennoch ist ein einheitlicher Bezugsrahmen zur Vergleichbarkeit der Werte nötig. Wie bereits oben gesagt, sind die im SETK 3-5 in der Auswertung ermittelten T-Werte nicht für die Analyse geeignet, da sie zum einen nicht separat für die sinnvollen Sätze der Aufgabe "Satzgedächtnis" und auch nicht für die

12 Monate wurden ins Dezimalsystem übertragen, um die Berechnungen zu erleichtern.

Aufgabe "Nachsprechen von Sätzen" aus Delfin 4 vorliegen und zum anderen die Testschwierigkeit der Untertests nicht wiedergeben und somit für den Vergleich der Testschwierigkeit zwischen den Gruppen nicht geeignet sind.

Um den erreichten Rohwert in Bezug zur maximal erreichbaren Punktzahl zu setzen und somit den Aufgaben einen einheitlichen Bezugsrahmen zu geben, wird für die Analyse der Prozentwert der maximal erreichbaren Punktzahl berechnet. Hat ein Kind also beispielsweise in Untertest 4 (Höchstpunktzahl 36) 18 Punkte erreicht, so liegt seine Leistung bei 50%. Die Umrechnung in Prozentwerte soll einen einheitlichen Bezugsrahmen für die Rohwerte gewährleisten, ohne dass auf die muttersprachlichen Normen der T-Werte zurückgegriffen werden muss. Für die statistische Analyse wurde der gemittelte Prozentwert der Untertests 1, 3 und 4 herangezogen, um sich auch beim ausschnittweisen Einsatz des SETK 3-5 an ein möglichst globales Maß der Sprachkompetenz anzunähern und damit eine aussagekräftige Einschätzung derselben zu treffen.

6.2 Ergebnisse

Die Kinder mit DaM schnitten durchschnittlich besser in den Aufgaben 1, 3 und 4 des SETK 3-5 ab ($M = 65,34\%$; $SD = 9,06$) als die Kinder mit DaZ ($M = 51,89\%$; $SD = 15,76$). Der Leistungsabstand vergrößerte sich jedoch im Untertest 2 „Satzgedächtnis“, Teil 1 mit einer Leistung von $56,74\%$ ($SD = 21,99$) der Kinder mit DaZ und $80,43\%$ ($SD = 11,11$) der Kinder mit DaM. Auch in der Aufgabe „Sätze nachsprechen“ aus Delfin 4 unterschieden sich die beiden Gruppen mit durchschnittlich $94,24\%$ ($SD = 12,89$) Leistung der Kinder mit DaM und 60% ($SD = 31,27$) Leistung der Kinder mit DaZ. Bereits bei Betrachtung dieser Werte fällt auf, dass beide Sprachgruppen verhältnismäßig hohe Werte in der Aufgabe "Sätze nachsprechen" erzielen. Dies ist ein Indikator dafür, dass die Aufgabe "Sätze nachsprechen" eine niedrigere Schwierigkeit aufweist, als andere Untertests des SETK 3-5. Tabelle 2 zeigt die Leistung in Prozent der erreichbaren Punktzahl in den beiden Sprachgruppen für die einzelnen Untertests.

Tabelle 2: Leistung der Kinder mit DaM und DaZ in den Untertests des SETK 3-5 und Delfin 4 "Sätze nachsprechen".

Untertest	DaM	DaZ
Verstehen von Sätzen	71,52 % (SD = 10,92)	54,00 % (SD = 20,48)
Satzgedächtnis (Teil 1)	80,43 % (SD = 11,11)	56,74 % (SD = 21,99)
Phonologisches Gedächtnis für Nichtwörter	60,61 % (SD = 22,02)	56,67 % (SD = 29,61)
Morphologische Regelbildung	63,64 % (SD = 14,49)	45,00 % (SD = 14,15)
Wiederholen von Wörtern (Anzahl der Wörter)	3,41 (SD = 0,50)	3,56 (SD = 0,53)
"Sätze nachsprechen" aus Delfin 4	94,24 % (SD = 12,90)	60,00 % (SD = 31,27)

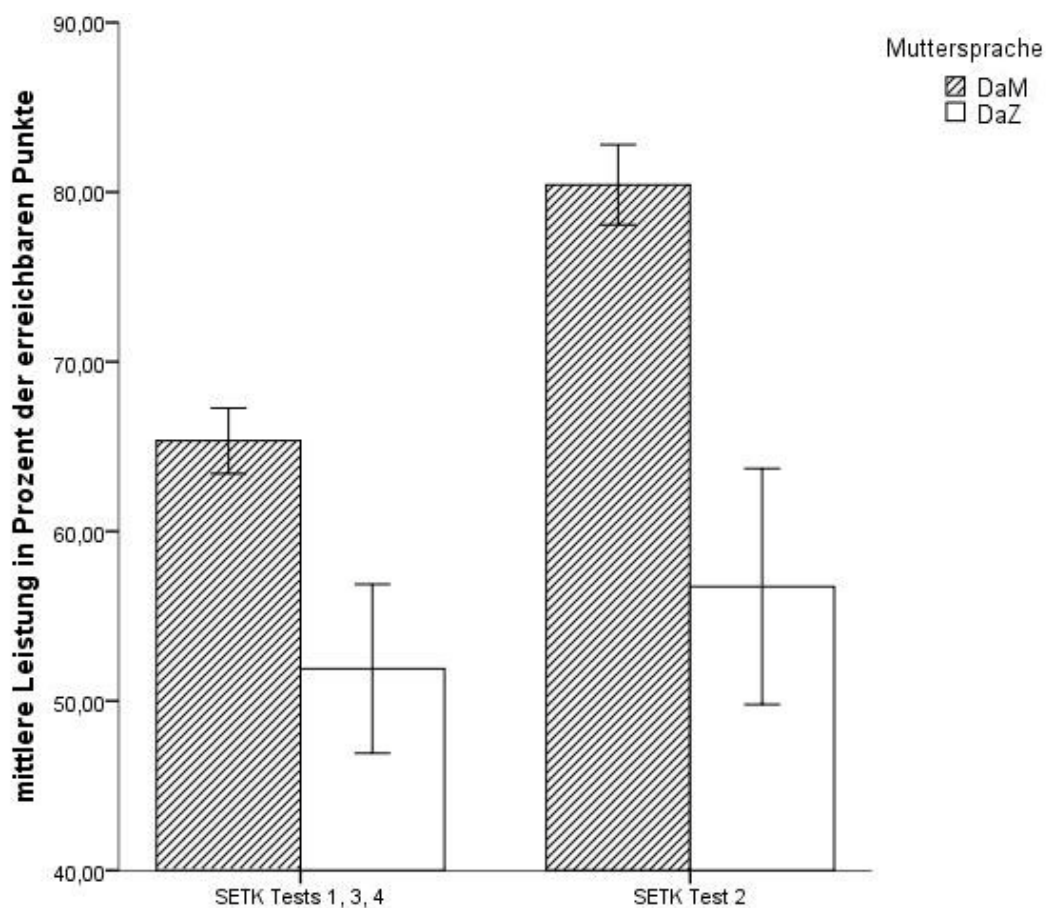


Abbildung 5: Vergleich der mittleren Leistung (als Prozent der maximal erreichbaren Punkte) von Vierjährigen mit DaM und DaZ für die Untertests 1, 3 und 4 des SETK 3-5 und den Untertest 2 "Satzgedächtnis".

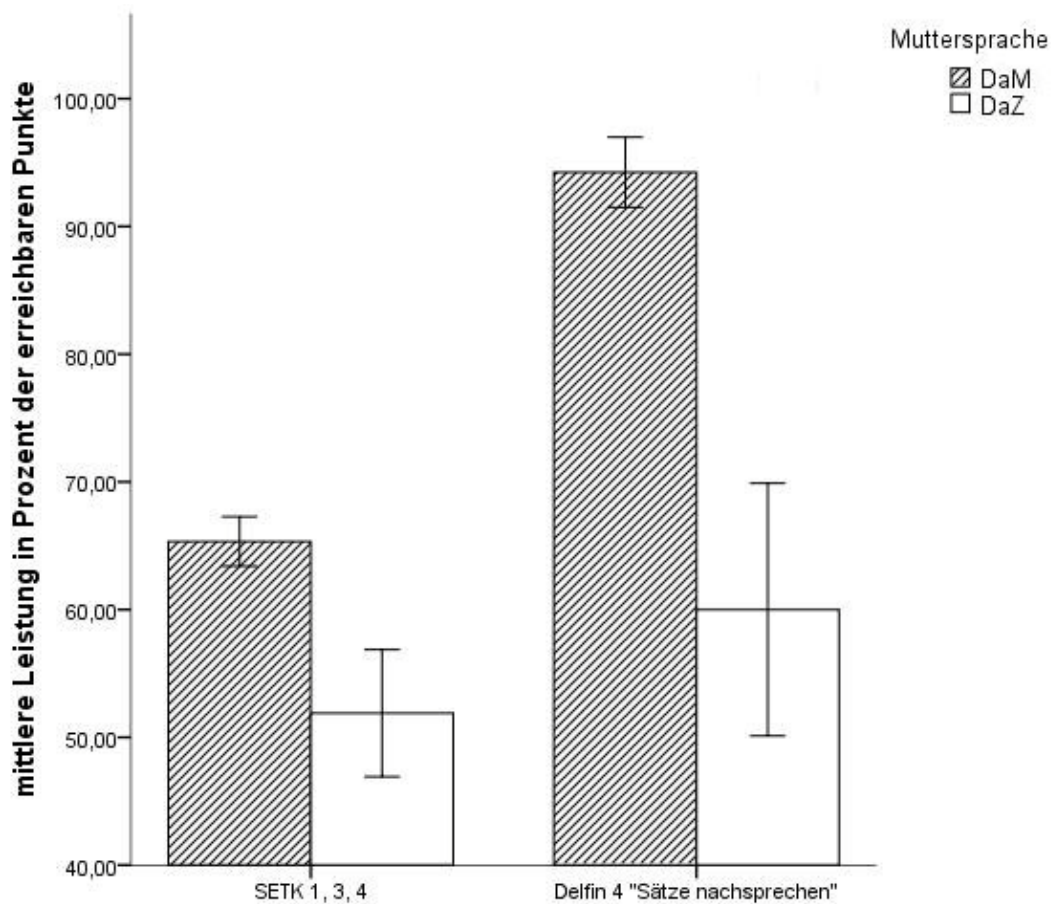


Abbildung 6: Vergleich der mittleren Leistung (als Prozent der maximal erreichbaren Punkte) von Vierjährigen mit DaM und DaZ für die Untertests 1, 3 und 4 des SETK 3-5 und der Aufgabe "Sätze nachsprechen" aus Delfin 4.

Eine 2x2 ANOVA mit Messwiederholung mit den Faktoren L1 (DaZ – DaM) und TEST (SETK 3-5 Untertests 1, 3, 4 – SETK 3-5 Untertest 2, Teil 1) zeigte durch einen Haupteffekt für L1 einen Unterschied zwischen den beiden Sprachgruppen auf mit $F(1,30) = 16,44$; $p < .001$ und einen Haupteffekt für den Faktor TEST mit $F(1,30) = 18,26$; $p < .001$. Die Faktoren L1 und TEST interagieren mit $F(1,30) = 4,8$; $p = .036$, aufgrund schlechterer Leistung der Kinder mit DaZ im Untertest „Satzgedächtnis“ (Teil 1) des SETK 3-5 (vgl. Abbildung 5). Geplante Einzelvergleiche zeigten signifikante Unterschiede zwischen den Gruppen für das kombinierte Maß der Untertests 1, 3 und 4 ($F(1,30) = 9,425$; $p = .005$) und den Aufgabentyp "Sätze nachsprechen" aus dem SETK 3-5 ($F(1,30) = 16,682$; $p < .001$).

Im Hinblick auf den Delfin 4 Untertest "Sätze nachsprechen" stellen sich die Ergebnisse ähnlich dar. In einer 2x2 ANOVA mit Messwiederholung und den Faktoren L1 (DaZ – DaM) und TEST (SETK 3-5 Untertests 1, 3, 4 – Delfin 4 „Sätze nachsprechen“) zeigen sich wiederum Haupteffekte für L1 ($F(1,30) = 19,81$; $p < .001$) und TEST ($F(1,30)$

= 32,03; $p < .001$). Weiterhin findet sich auch hier eine Interaktion der Faktoren L1 und TEST mit $F(1,30) = 10,12$; $p = .003$ (vgl. Abbildung 6). Ein Einzelvergleich zwischen den Gruppen für die Leistung in der Aufgabe "Sätze nachsprechen" aus Delfin 4 war hochsignifikant ($F(1,30) = 19,673$; $p < .001$).

Da mit Blick auf den Altersunterschied zwischen den beiden Sprachgruppen (der oben bereits berichtet wurde) die Interpretation dieser Gruppenunterschiede problematisch erscheint, wurden außerdem Kovarianzanalysen mit der Kovariate ALTER gerechnet. Dabei zeigte sich, dass die Interaktion zwischen den Faktoren L1 und TEST beim Vergleich SETK 3-5 1, 3, 4 und Delfin 4 „Sätze nachsprechen“ signifikant bleibt mit $F(1,29) = 6,31$; $p = .018$. (Haupteffekt für L1: $F(1,29) = 13,88$; $p = .001$; kein Haupteffekt für TEST: $F < 1$).

Beim Vergleich SETK 3-5 1, 3, 4 und SETK 3-5 2, Teil 1 mit der Kovariate ALTER zeigt sich allerdings keine Interaktion der Faktoren L1 und TEST: $F(1,29) = 2,29$; $p = .14$ (Haupteffekt für L1: $F(1,29) = 11,15$; $p = .002$; kein Haupteffekt für Test: $F < 1$).

Tabelle 3: Standardisierte Regressionskoeffizienten und Signifikanz für multiple Regression, abhängige Variable: Leistung in der Satzwiedergabe (SETK 3-5, Untertest 2, Teil 1).

Variable	β	Sign. ^a
Leistung SETK 3-5, Untertests 1, 3, 4	.561	**
L1 ^b	.323	*
Alter	.180	
R^2	.773	

a) * $p < .05$; ** $p < .01$

b) Dummy-Variable: 0 = DaZ; 1 = DaM

Im Folgenden soll der Zusammenhang zwischen der Leistung in der Aufgabe "Sätze nachsprechen" und den Prädiktoren Alter, L1 und Testergebnis in den SETK 3-5 Untertests 1, 3 und 4 genauer betrachtet werden. Dazu wird eine lineare multiple Regression mit der abhängigen Variable "Sätze nachsprechen" (SETK Untertest 2, Teil 1) und den Prädiktoren SETK 3-5 (Untertests 1, 3, 4), Alter und L1 gerechnet. Dabei zeigte sich die Leistung in den Untertests 1, 3 und 4 als bester Prädiktor für die Leistung im

"Sätze nachsprechen" mit $\beta = .561$ ($p < .001$). Auch für den Faktor L1 wurde ein signifikanter Einfluss auf die Leistung in der Satzwiedergabe gefunden mit $\beta = .323$ ($p = .023$). Für den Einfluss des Alters auf das Ergebnis in der Aufgabe "Satzwiedergabe" wurde hier kein Hinweis gefunden ($\beta = .18$; $p = 0,839$). Die Ergebnisse sind in Tabelle 3 als Übersicht dargestellt.

6.3 Diskussion

Wie erwartet, unterscheiden sich die Kinder mit DaM und DaZ in ihrer Leistung in allen Aufgabentypen (Haupteffekt für L1 in allen Analysen). Dies kann damit erklärt werden, dass die DaZ-Kinder noch nicht lange und intensiv genug Kontakt zu ihrer L2 hatten, um in ihrer Leistung zur Gruppe der Muttersprachler aufzuschließen. Der in den Analysen auftretende Haupteffekt für TEST lässt darauf schließen, dass die Testschwierigkeit der Aufgabe „Sätze nachsprechen“ (insbesondere für Kinder mit DaM) niedriger ist als die der übrigen Aufgabentypen, so dass auch im Verhältnis zur maximalen Punktzahl leichter höhere Werte erreicht werden können.

Das für diese Studie zentrale Ergebnis ist das aufgabenspezifisch schlechtere Abschneiden der Kinder mit Deutsch als Zweitsprache bei der Satzwiedergabe, das sich in der Interaktion zwischen TEST und L1 zeigt: Kinder mit DaZ schnitten im Aufgabentyp „Sätze nachsprechen“ in beiden Vergleichen (SETK 3-5 Untertest 2 und Delfin 4 „Sätze nachsprechen“) schlechter ab als anhand ihrer Leistungen aus den übrigen SETK 3-5 Aufgaben und der Leistung der Kinder mit DaM zu erwarten wäre. Dies ist Evidenz dafür, dass Kinder mit DaZ durch diesen Aufgabentyp benachteiligt werden.

Das Ergebnis der zweiten Kovarianzanalyse (Vergleich SETK 3-5, Untertests 1, 3, 4 und SETK 3-5, Untertest 2 mit Kovariate Alter) ist problematisch, da es eine Alternativerklärung der Ergebnisse nahelegt. Die Interaktion zwischen L1 und TEST in der oben berichteten Analyse könnte nicht auf die Erstsprache (DaM oder DaZ) der Kinder zurückzuführen sein, sondern auf ihr Alter, da die Probandengruppe mit DaM gleichzeitig auch die ältere Probandengruppe ist. In diesem Fall wären es nicht Kinder mit DaZ, sondern jüngere Kinder, die durch diese Aufgabe nachteilig bewertet werden würden. Es wäre also möglich, dass das schlechte Abschneiden in der Aufgabe "Satzgedächtnis" des SETK 3-5 im Vergleich zur Leistung in den anderen Testaufgaben auf das Alter und nicht auf die L1 zurückzuführen ist. Auch der Haupteffekt zwischen den beiden Gruppen könnte auf den Altersunterschied zurückgeführt werden statt auch den Unterschied in der

Erstsprache. Gegen diese Interpretation spricht zum einen die Kovarianzanalyse, die die Aufgabe "Sätze nachsprechen" aus Delfin 4 mit den Untertests des SETK 3-5 vergleicht: Hier bleibt die Interaktion signifikant, auch wenn das Alter als Kovariate berücksichtigt wird. Daneben gibt es keine theoretische Annahme, die die Interpretation nahelegt, dass das Alter die Leistung in der Aufgabe "Satzgedächtnis" in anderem Maße beeinflusst als die Leistung in den anderen Testaufgaben. Drittens spricht die Regressionsanalyse, die keinen systematischen Einfluss des Alters auf die Leistung in der Satzwiedergabe findet, aber wohl einen signifikanten Einfluss der Muttersprache gegen diese Annahme. Dennoch soll diese Interpretation durch eine weitere Analyse geprüft werden, deren Logik im Folgenden beschrieben wird.

Wenn die Unterschiede zwischen den Sprachgruppen auf deren Altersunterschied zurückzuführen sind, so würde daraus geschlossen werden, dass jüngere Kinder in der Aufgabe "Satzgedächtnis" schlechter abschneiden als aus ihrer Leistung in den übrigen Tests und der Leistung älterer Kinder erwartet werden würde. Dies sollte sich in einer Interaktion der Faktoren TEST und ALTER niederschlagen.

6.3.1 Einfluss des Alters auf die Aufgabe "Sätze nachsprechen" in der L1

Um den Einfluss der Variable ALTER von dem der Variable L1 zu trennen, wurden für die folgende Analyse nur die Kinder mit DaM betrachtet. Diese wurden entsprechend ihres Alters in drei Gruppen eingeteilt, dabei waren sechs Kinder in Gruppe 1 (4;00 - 4;05 Jahre), davon ein Junge und fünf Mädchen, elf Kinder in Gruppe 2 (4;06 – 4;11 Jahre), davon sechs Jungen und fünf Mädchen und fünf Kinder in Gruppe 3 (5;00 – 5;05), alle davon Jungen.

In einer ANOVA mit Messwiederholung und den Faktoren ALTER (Gruppe 1 – Gruppe 2 – Gruppe 3) und TEST (SETK 3-5 Untertests 1, 3, 4 – SETK 3-5 Untertest 2, Teil 1) zeigte sich ein Haupteffekt für TEST mit $F(1,19) = 46,29; p < .001$, aber kein Haupteffekt für die Altersgruppe ($F < 0,2$). Es zeigte sich außerdem keine Interaktion zwischen den Faktoren ALTER und TEST für die Gruppe der DaM-Kinder ($F(2,19) = 1,32; p > .25$).

Diese nachträgliche Analyse lässt den Schluss zu, dass zumindest in der Gruppe der Muttersprachler das Alter keinen größeren Einfluss auf die Leistung in der Aufgabe

"Satzwiedergabe" hat als auf die Leistung in den anderen Aufgaben. Obwohl die Konfundierung von L1 und ALTER in den Gruppen problematisch ist, ist es wahrscheinlich, dass die Unterschiede zwischen den Gruppen auf die Unterschiede in der Erstsprache (DaZ und DaM) und nicht auf die zwischen den Sprachgruppen bestehenden Altersunterschiede zurückzuführen sind.

Eine weitere Einschränkung der Aussagekraft dieser Studie besteht durch den allgemeinen Leistungsunterschied zwischen den beiden Gruppen. Selbst wenn der Leistungsunterschied in der Aufgabe "Satzgedächtnis" bzw. "Sätze nachsprechen" auf das Testen in einer Zweitsprache zurückgeführt werden kann, so wäre es denkbar, dass sich die Leistung der Kinder mit DaZ mit dem fortschreitenden Spracherwerb auch in diesem Aufgabentyp an die Leistung der Muttersprachler angleicht. Nach dem weitestgehend vollständigen Erwerb der L2 sollte dann kein Unterschied zwischen Muttersprachlern und hochkompetenten Sprechern einer L2 (*Near-Natives*) mehr nachweisbar sein. Dies würde darauf hindeuten, dass die Unterschiede in der Leistung in der Satz-wiedergabe tatsächlich mit dem noch nicht vollständigen Spracherwerb in der L2 zusammenhängen und damit zumindest bedingt Aussagekraft über die Sprachkompetenz der Kinder haben. Hier soll jedoch gezeigt werden, dass diese Unterschiede mit der Verarbeitung einer Zweitsprache auch auf höchstem Niveau zusammenhängen. Es handelt sich um generelle Verarbeitungsunterschiede für die L2, die auch über die Erwerbsphase hinaus bestehen. Um zu überprüfen, ob dies der Fall ist, wurden in Studie 2 erwachsene Muttersprachler und *Near-Natives* mit Deutsch als L2 untersucht. Der Vergleich von erwachsenen *Near-Natives* mit abgeschlossenem Spracherwerb und Muttersprachlern soll zeigen, dass die Verarbeitungsunterschiede auch nach erfolgreichem L2-Erwerb weiter bestehen.

7 Studie 2 - Satzbehalten bei Erwachsenen mit L1 / L2 Deutsch

In dieser zweiten Studie soll überprüft werden, ob sich erwachsene *Near-Natives* in der Aufgabe „Sätze wiederholen“¹³ von Muttersprachlern unterscheiden. Dazu werden in einem Parallelisierungsverfahren Muttersprachler des Deutschen und *Near-Natives*, die Deutsch auf muttersprachlichem Niveau sprechen, anhand ihrer Leistung in einem formalen Sprachtest parallelisiert. Es wird erwartet, dass *Near-Natives*, die sich in ihrer Sprachkompetenz nicht von Muttersprachlern unterscheiden, in der Satzwiedergabe schlechter abschneiden.

Weiterhin soll ausgeschlossen werden, dass dieser Unterschied darauf zurückzuführen ist, dass die Satzwiedergabe ein sensiblerer Indikator der Sprachkompetenz ist als der eingesetzte alternative Sprachtest. Dies würde bedeuten, dass die unterschiedliche Leistung in Satzwiedergabe und alternativem Sprachtest lediglich daher rührte, dass die Satzwiedergabe auch zwischen Leistungsunterschieden differenziert, die der alternative Test nicht abbilden kann.

Um diese Erklärung auszuschließen, muss gezeigt werden, dass die Satzwiedergabe tatsächlich andere Kompetenzen bzw. Komponenten der Sprachverarbeitung erfasst als der alternative Sprachtest (im Kontrast zu der Annahme, dass er die gleichen erfasst, aber auf einer stärker differenzierenden Skala). Dazu werden im zweiten Schritt dieser Studie *Near-Natives* und Muttersprachler anhand ihrer Leistung in der Satzwiedergabe parallelisiert. Dabei wird erwartet, dass Muttersprachler, die auf dem Leistungsniveau von *Near-Natives* in der Satzwiedergabe liegen, schlechtere Sprachkompetenz im alternativen Sprachtest aufzeigen. Durch diese zweifache Dissoziation der beiden Maße soll aufgezeigt werden, dass tatsächlich unterschiedliche Fähigkeiten die gemessene Leistung in unterschiedlichem Maße beeinflussen.

¹³ Da in den Studien mit älteren Teilnehmern die Wiedergabe der Sätze schriftlich statt mündlich erfolgt, wird die Bezeichnung "Sätze nachsprechen" durch die allgemeinere Bezeichnung "Sätze wiederholen" bzw. Satzwiedergabe ersetzt. Die theoretischen Vorannahmen sind für beide Fälle gleichermaßen gültig.

7.1 Methode

Stichprobe

24 Studenten (18 davon weiblich) der Universität Erfurt mit Deutsch als Muttersprache und einem durchschnittlichen Alter von 21,4 Jahren ($SD = 1,88$, Range: 18 - 25 Jahre) und zwölf *Near-Natives* (elf davon weiblich) mit einem durchschnittlichen Alter von 24,42 Jahren ($SD = 5,18$; Range: 18 - 37 Jahre), die entweder Studenten oder Mitarbeiter an Universitäten in Erfurt, Hamburg oder Saarbrücken waren, nahmen gegen eine Aufwandsentschädigung oder Versuchspersonenstunden an der Studie teil. Die Daten dieser beiden Gruppen konnten hinsichtlich ihrer Leistung im C-Test parallelisiert werden, aber nicht hinsichtlich ihrer Leistung in der Satz wiedergabe. Zur Parallelisierung mit der Gruppe der *Near-Natives* in der Satz wiedergabeleistung wurden die Daten von 21 Schülern eines Gymnasiums in Erfurt, 13 davon Mädchen, erhoben. Alle Schüler sprachen Deutsch als Muttersprache, standen am Ende ihres 7. Schuljahres und nahmen gegen eine Aufwandsentschädigung an der Studie teil. Sie waren zum Zeitpunkt der Erhebung zwischen 13 und 14 Jahre alt, nur ein Schüler hatte das 14. Lebensjahr bereits vollendet.

Alle *Near-Natives* sprachen akzentfreies Deutsch und lebten seit mindestens 5 Jahren in Deutschland ($M = 9,46$ Jahre, $SD = 4,05$; Range: 5 - 19 Jahre). Zehn von ihnen hatten in Deutschland das Abitur absolviert, alle studierten oder promovierten an einer deutschen Hochschule. Im Mittel hatten sie in einem Alter von 13,67 ($SD = 6,54$; Range: 7 - 27¹⁴) begonnen, Deutsch zu lernen. Ihr langer Aufenthalt in Deutschland und ihre erfolgreiche Bildungsgeschichte sprechen für eine sehr hohe Kompetenz in der deutschen Sprache. Die Muttersprachen der *Near-Natives* waren Russisch (7 TN), Englisch (2 TN), Polnisch (1 TN), Französisch (1 TN) und Rumänisch (1 TN).

Material

C-Test

Um die Sprachkompetenz der *Near-Natives* formal zu erfassen, wurde ein Sprachtest nach dem C-Test-Prinzip von Raatz und Klein-Braley (1982) eingesetzt. Dieses ökonomische Testverfahren ist einer der wenigen Sprachtests, die auch für den Einsatz mit erwachsenen und nicht beeinträchtigten Muttersprachlern geeignet sind. Das Prinzip des C-

¹⁴ Dabei lag das AoA von zehn der zwölf *Near-Natives* zwischen 7 und 18 Jahren; zwei Teilnehmerinnen hatten erst im Alter von 25 und 27 begonnen, ihre L2 Deutsch zu erwerben. Dennoch war auch bei diesen Sprecherinnen kein Akzent während eines natürlichen Gesprächs erkennbar.

Tests baut auf der Redundanz von Sprache auf, also darauf, dass sprachliche Information gewöhnlich reichhaltiger ist, als zu ihrem Verständnis nötig ist. Dies stellt sicher, dass eine Botschaft auch unter Störeinflüssen decodiert werden kann. Solche Störeinflüsse können z.B. Umgebungslaute, geteilte Aufmerksamkeit, begrenzte Merkfähigkeit sein; bei der Schriftsprache können sie durch undeutliche Schrift oder fehlende Teile entstehen. Im C-Test wird ein systematischer Störeinfluss geschaffen, indem Wortteile systematisch gelöscht werden: Ab dem zweiten Satz wird jeweils die zweite Hälfte jedes zweiten Wortes entfernt. Diese Lücken können durch die redundante Information, die weitgehend automatisch verarbeitet wird, gefüllt werden, und die Botschaft kann wieder hergestellt werden. Dazu ist eine hohe allgemeine Sprachkompetenz nötig, die die Vertrautheit mit Struktur und Lexikon der Sprache voraussetzt. Da der C-Test zumeist (wie auch in dieser Studie) unter Zeitdruck angewandt wird, erfasst er teilweise auch den Automatisierungsgrad der Verarbeitung. So würde sich ein relevanter niedriger Automatisierungsgrad in einer längeren Bearbeitungsdauer niederschlagen und damit zu einer geringeren Punktzahl führen.

Das C-Test-Prinzip (Raatz & Klein-Braley, 1982) hat sich in zahlreichen Studien als valide und reliabel erwiesen, seine Objektivität ist durch das Auswertungsschema ebenfalls gegeben. Einen Überblick über die Forschung zum C-Test gibt Grotjahns C-Test Bibliographie (2007). Der C-Test ist eines der wenigen Instrumente, die für Tests in einer L1 und L2 geeignet sind (vgl. Grotjahns C-Test Bibliographie, 2007). Da er als Maß der allgemeinen Sprachkompetenz eine zentrale Rolle in dieser Arbeit einnimmt, sollen hier Belege für die Validität seines Konstruktionsprinzips aufgeführt werden.

C-Test-Ergebnisse von jungen deutschen Muttersprachlern korrelieren mit der Klassenstufe, Schulform (Gymnasium, Realschule, Hauptschule) und der Note im Fach Deutsch (Wockenfuß & Raatz, 2006). Für Zweitsprachler des Deutschen finden Studien im Auftrag des TestDaF-Instituts Korrelationen zwischen C-Tests und dem "Test Deutsch als Fremdsprache" (TestDaF), einem dreistündigen, umfassenden Testverfahren, das in Deutschland als offizieller Nachweis der Sprachkompetenz gegenüber Universitäten und Arbeitgebern etabliert ist (vgl. Arras, Eckes & Grotjahn, 2002; Eckes, 2010). Der dort verwendete C-Test korreliert mit den einzelnen Untertests des TestDaF (Leseverstehen, Hörverstehen, schriftlicher Ausdruck und mündlicher Ausdruck), die höchste Korrelation besteht jedoch zwischen dem C-Test und dem Gesamtergebnis des TestDaF ($r = 0,76$; $p < .01$; $N = 470$), was darauf hinweist, dass der C-Test am ehesten die allgemeine

Sprachfähigkeit abbildet (Eckes, 2010). Für englische C-Tests wurden signifikante Korrelationen mit dem "Test of English as a Foreign Language" (TOEFL; $r = 0,55$) und dem "Test of English International Communication" (TOEIC; $r = 0,62$) gefunden (Eckes, 2010). Für erwachsene Muttersprachler mit hohem Bildungsgrad sind in den meisten C-Tests sehr hohe Werte zu erwarten. Dennoch ist es möglich, C-Tests so konstruieren, dass auch Muttersprachler nicht die volle Punktzahl erreichen (Baur & Meder, 1994). Der in den folgenden Studien eingesetzte C-Test wird (zum Teil mit leichten Variationen) an verschiedenen Hochschulen und Sprachinstituten verwendet (z.B. Institut für Internationale Kommunikation (IIK), Düsseldorf) und ist so konstruiert, dass auch Muttersprachler selten die volle Punktzahl erreichen. Er besteht aus 4 Kurztexten mit insgesamt 99 Lücken¹⁵. Die Kurztexte variieren thematisch und sind entsprechend ihrer Schwierigkeit aufsteigend geordnet (siehe Anhang A2). Für den C-Test wurde ein Zeitlimit von 20 Minuten (5 Minuten pro Kurztext) gesetzt.

Sätze

Für die Satzwiedergabe wurden 40 der längsten Sätze aus der deutschen Version des Lesespannentests von Daneman und Carpenter (1980) ausgewählt (Hacker, Handrick & Veres, 1996) und um einzelne Wörter ergänzt, so dass die Sätze länger wurden. Dies war nötig, um Deckeneffekte in der Leistung der Muttersprachler zu vermeiden. Die jeweiligen Strukturen und der allgemeine Inhalt der Sätze blieben durch die Ergänzungen unverändert (siehe Beispiel 2). So entstanden 40 Sätze mit einer Länge von 16-20 Wörtern ($M = 17,8$, $SD = 1,12$). Diese wurden mit der Methode der Lateinischen Quadrate auf 4 Listen aufgeteilt, die jeweils mittlere Satzlängen von 17,9, 17,7, 18 und 17,5 hatten. Die Mehrzahl der Sätze war grammatikalisch komplex, bestand also aus Haupt- und Nebensätzen mit mehreren Nominalphrasen. Es waren jedoch keine ungewöhnlichen Strukturen (wie z.B. Holzwegsätze) im Material enthalten (siehe Anhang A1). Da diese Studie zum Ziel hat, einen materialübergreifenden Effekt aufzuzeigen, wurde der Einfluss, den einzelne syntaktische Strukturen auf die Satzwiedergabe haben, nicht berücksichtigt.

Die Darbietung der Sätze erfolgte auditiv. Die Wiedergabe der Sätze erfolgte schriftlich auf vorbereiteten Antwortbögen¹⁶. Die Sätze wurden von einer

¹⁵ Im Original des Tests finden sich 100 Lücken; eine davon war fehlerhaft gesetzt in dem Sinne, dass sie nicht dem C-Test-Prinzip entsprach. Für diese Arbeit wurde dies korrigiert, was dazu führte, dass der hier verwendete C-Test nur noch 99 Lücken enthält.

¹⁶ Um auszuschließen, dass die Ergebnisse auf die Nutzung schriftlicher Wiedergabe zurückzuführen sind, wurde eine Kontrollstudie mit mündlicher Wiedergabe durchgeführt. Hier fanden sich vergleichbare Ergebnisse (Schmitterer, Knackstedt, Salomon, Scholze & Schweppe, 2009, unveröffentlichte Daten).

Muttersprachlerin des Deutschen aufgesprochen. Sie wurden über Kopfhörer wiedergegeben, die einzelnen Audiodateien hatten eine Länge von 6-8 Sekunden.

Original	Ganz besonders liebe ich stille Abende zu zweit, mit Wein, Musik, und dem Schein der Kerze.
----------	---

Adaptiert	Ganz besonders liebe ich stille Abende zu zweit, mit Wein, Musik, gutem Essen und dem Schein einer
-----------	--

Original	Die Spur des Mädchens führte zu einem sehr großen Felsblock und verlor sich an dieser Stelle.
----------	---

Adaptiert	Die Spur des Mädchens führte zu einem sehr großen Felsblock und verlor sich an dieser Stelle im Nichts.
-----------	---

Beispiel 2: Originalsätze aus dem Lesespannentest und adaptierte Sätze für Studie 2

In dieser Studie wird die Leistung im C-Test mit der Leistung in der Satzwiedergabe verglichen. Dieser Vergleich bietet sich an, da beide als Tests der allgemeinen Sprachkompetenz angelegt sind, sich aber in ihrer Beanspruchung von Aufmerksamkeit und Arbeitsgedächtnis unterscheiden. Während die Satzwiedergabe hohe Anforderungen an diese Ressourcen stellt, belastet der C-Test sie nur minimal. Dies ist ein entscheidender Unterschied, da nach den oben diskutierten Modellen der Satzverarbeitung die Unterschiede zwischen Muttersprachlern und Zweitsprachlern dann evident werden, wenn die Ressource Aufmerksamkeit über ihre Grenzen hinweg beansprucht wird. Abseits dieses Unterschieds beziehen C-Test und Satzwiedergabe sowohl das Sprachverständnis als auch die Sprachproduktion mit ein; Kenntnisse auf allen Verarbeitungsebenen (Phonologie, Morphologie, Syntax, lexikalisches Wissen) sind nötig für die Bearbeitung. Desweiteren arbeiten beide Aufgabentypen mit vorgegebenen Formulierungen, es ist den Teilnehmern also nicht möglich, auf alternative, unter Umständen besser beherrschte, grammatikalische Konstruktionen und Wörter auszuweichen. Dieses Ausweichen wäre für die Zweitsprachler eine Möglichkeit, vorhandene Probleme mit ihrer L2 zu verbergen.

Ablauf

Die Teilnehmer bearbeiteten zuerst eine gedruckte Version des C-Tests, danach wurden ihnen am Computer Sätze präsentiert. Die Probanden wurden instruiert, mit der

Wiedergabe zu beginnen, sobald der Satz vollständig angehört wurde, aber keinesfalls früher. Die Sätze waren zwischen 6 und 8 Sekunden lang und in natürlicher Geschwindigkeit gesprochen. Die Wiedergabe erfolgte schriftlich auf vorbereiteten Antwortbögen und ohne Zeitdruck. Im Anschluss füllten die Teilnehmer einen Fragebogen zu ihrer Spracherwerbsgeschichte aus (siehe Anhang B2). Die Sitzungen dauerten insgesamt ca. 30 Minuten. Jedem Teilnehmer wurden jeweils 10 Sätzen präsentiert, die Satzlisten wurden balanciert über die Teilnehmer hinweg gezeigt. Zuvor durchliefen die Probanden zwei Übungsdurchgänge und hatten die Gelegenheit, Fragen zum Ablauf zu klären.

Auswertung

C-Test

Der C-Test wurde gemäß der Musterlösung ausgewertet, für jede korrekt gefüllte Lücke wurde ein Punkt vergeben, der höchste erreichbare Wert waren 99 Punkte. Die Gruppe der erwachsenen Muttersprachler erreichte im Durchschnitt einen Wert von 95,2 Punkten ($n = 24$, $SD = 2,8$), die Gruppe der *Near-Natives* einen Wert von 93,33 ($n = 12$, $SD = 3,5$), die Gruppe der jugendlichen Muttersprachler erreichte durchschnittlich 89,24 Punkte ($n = 21$, $SD = 3,86$).

Tabelle 4: Übersicht über die Leistung der beiden muttersprachlichen Gruppen und der *Near-Natives* in den beiden Aufgabentypen.¹⁷

	C-Test	Satzwiedergabe
Muttersprachler (Studenten)	M = 93,55 (SD = 3,86) $n = 11$	M = 90,756 (SD = 6,738) $n = 11$
<i>Near-Natives</i>	M = 93,73 (SD = 3,73) $n = 11$ M = 93,417 (SD = 3,37) $n = 12$	M = 77,4 (SD = 8,562) $n = 11$ M = 76,515 (SD = 8,721) $n = 12$
Muttersprachler (Schüler)	M = 87,00 (SD = 3,075) $n = 12$	M = 76,363 (SD = 13,188) $n = 12$

Satzwiedergabe

In der Auswertung der Satzwiedergabe wurden für jeden Satz die korrekt wiedergegebenen Wörter gezählt. Dabei wurden nur genau wiedergegebene Wortformen akzeptiert, die Reihenfolge wurde nicht berücksichtigt. Dieses Auswertungsverfahren

¹⁷ Die Werte für die *Near-Natives* sind für die beiden Analysen separat angegeben, da in der ersten Analyse eine Teilnehmerin aufgrund des Matchingverfahrens ausgeschlossen wurde, da ihr C-Test-Ergebnis nicht exakt genug mit einem der noch nicht gematchten Muttersprachler gematcht werden konnte. Die Daten von 11 Teilnehmern wurden in beiden Analysen verwendet.

entspricht denen in Delfin 4 und dem SETK 3-5. Anschließend wurde für jeden Satz errechnet, wie viel Prozent der Wörter korrekt wiedergegeben wurden.

Parallelisierungsverfahren

Für die Analyse wurden die Stichproben anhand der Leistungskriterien in C-Test und Satzwiedergabe gepaart und ihre Leistung im jeweils anderen Kriterium wurde verglichen. Einen Überblick über die Mittelwerte der Gruppen, die in der Analyse berücksichtigt wurden, gibt Tabelle 4.

C-Test

Anhand des C-Tests wurde jedem der *Near-Natives* ein Muttersprachler zugeordnet, der mit gleicher Leistung abschnitt. Das Resultat waren 11 Paare¹⁸, von denen 8 perfekt gepaart werden konnten (exakt gleicher C-Test-Wert). Drei Paare unterschieden sich jeweils in einem Punkt, zwei davon zum Vorteil der *Near-Natives*. Der mittlere C-Test-Wert für die 11 Muttersprachler lag bei 93,55 (SD = 3,86), der mittlere C-Test-Wert der *Near-Natives* bei 93,73 (SD = 3,73).

Satzwiedergabe

Zur Parallelisierung der Stichproben wurde die Leistung in der Satzwiedergabe in Prozent verwendet. Dabei wurden 12 Paare mit je einem *Near-Native* und einem Muttersprachler gebildet, die einen vergleichbaren Prozentwert in der Wiedergabe erreicht hatten. Aufgrund der feineren Gradierung der Prozentwerte im Vergleich zu den C-Test-Punktwerten konnten keine numerisch gleichen Paare gebildet werden, die Abweichungen wurden jedoch so gering gehalten wie möglich, sie hatten einen durchschnittlichen Betrag von 3,59 % (SD = 2,08). Die Gruppe der *Near-Natives* zeigte eine durchschnittliche Wiedergabeleistung von 76,515 % (SD = 8,721), die Muttersprachler erreichten im Durchschnitt einen Wert von 76,363 % (SD = 13,188).

18 Die C-Test-Leistung einer Teilnehmerin der Near-Native-Gruppe konnte nicht ohne große Abweichung mit der Leistung eines Teilnehmers / einer Teilnehmerin aus der muttersprachlichen Gruppe gematcht werden. Daher werden diese Daten nicht in die Analyse mit einbezogen.

7.2 Ergebnisse

Parallelisierung nach C-Test

Zuerst soll nun die Leistung in der Satzwiedergabe verglichen werden für die elf Paare, die anhand ihrer Leistung im C-Test parallelisiert wurden. In einem T-Test für gepaarte Stichproben zeigte sich hier ein signifikanter Unterschied zwischen den beiden Gruppen in ihrer Wiedergabeleistung. Die *Near-Natives* ($M = 77,40\%$, $SD = 8,562$) schnitten schlechter ab als die Muttersprachler ($M = 90,756\%$, $SD = 6,738$) mit $t(10) = 4,818$; $p = .001$.

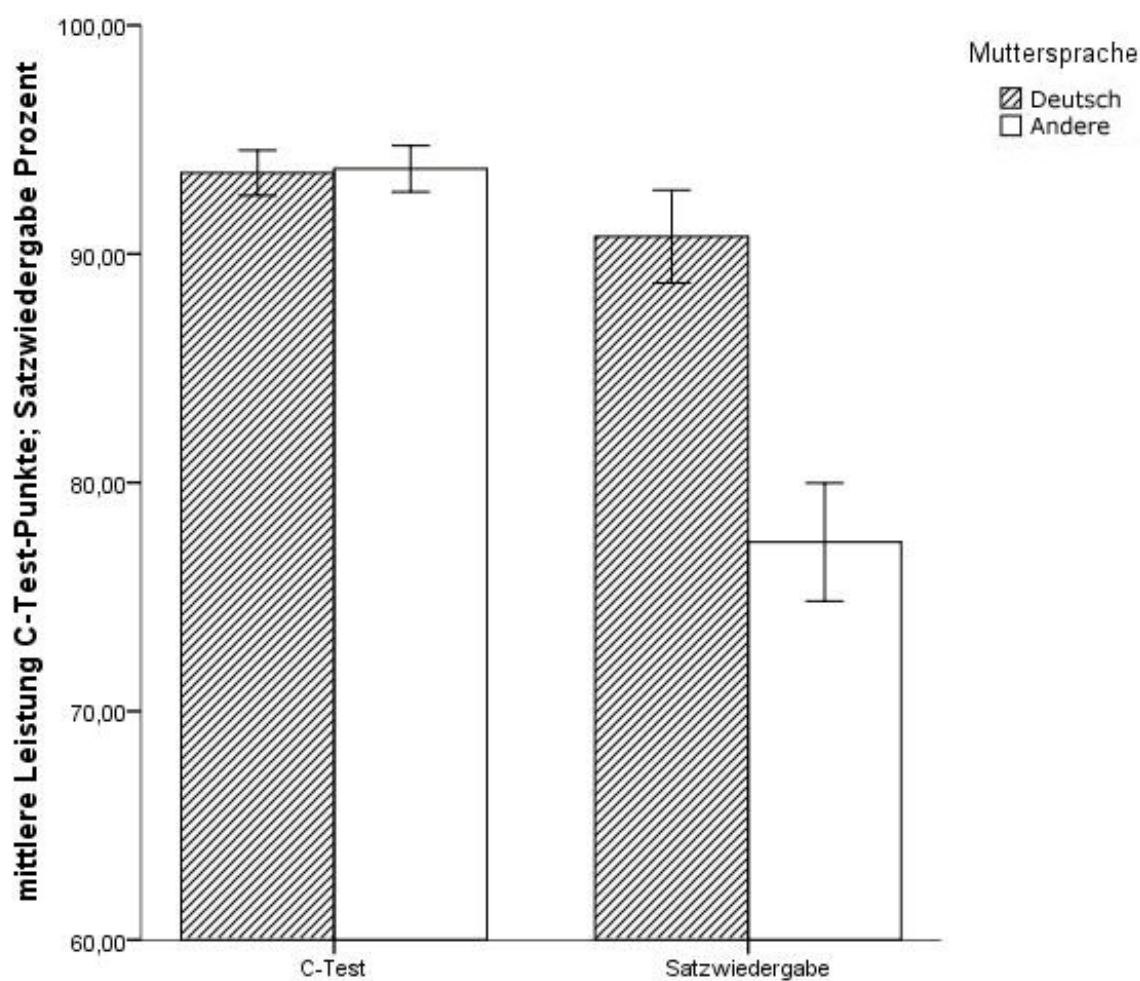


Abbildung 7: Vergleich der durchschnittlichen Leistung von Probanden mit Deutsch als L1 bzw. L2 im C-Test und in der Satzwiedergabe (als Prozent der korrekt wiedergegebenen Wörter pro Satz): Parallelisierung entsprechend C-Test-Leistung.

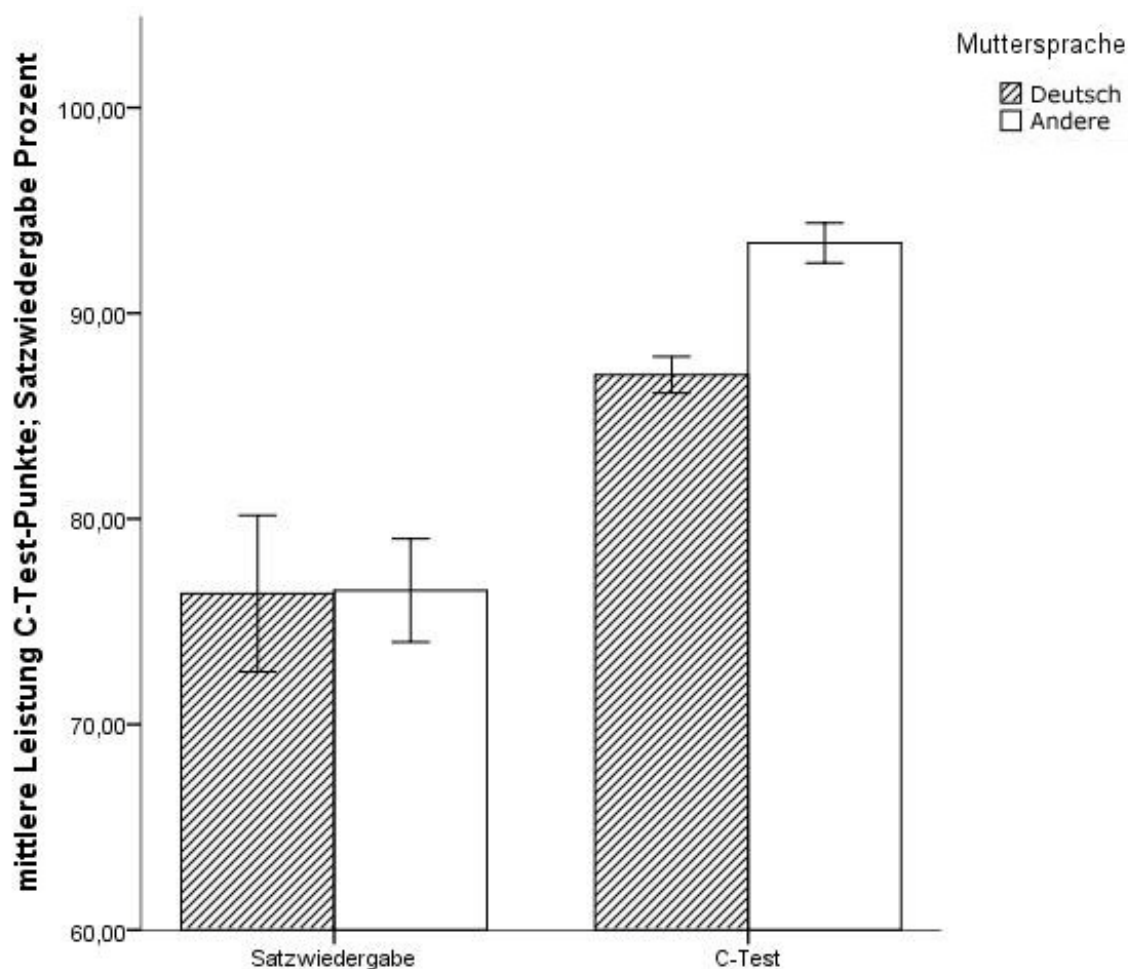


Abbildung 8: Vergleich der durchschnittlichen Leistung von Probanden mit Deutsch als L1 bzw. L2 im C-Test und in der Satzwiedergabe (als Prozent der korrekt wiedergegebenen Wörter pro Satz): Parallelisierung entsprechend Leistung in der Satzwiedergabe.

Parallelisierung nach Satzwiedergabe

In dieser Analyse soll nun die Leistung im C-Test verglichen werden für 12 Paare, die anhand ihrer Leistung in der Satzwiedergabe gematcht wurden. In einem T-Test für gepaarte Stichproben unterschieden sich die *Near-Natives* und die jugendlichen Muttersprachler in ihren C-Test-Ergebnissen. Die Muttersprachler ($M = 87,5$, $SD = 2,97$) schnitten durchschnittlich schlechter ab als die *Near-Natives* ($M = 93,4$, $SD = 3,37$) mit $t(11) = 5,64$; $p < .001$.

7.3 Diskussion

Auch *Near-Natives*, also hochkompetente Sprecher einer L2, schneiden bei der Wiedergabe von Sätzen schlechter ab als Muttersprachler. Sogar wenn ihre

Sprachkompetenz in ihrer L2 Deutsch durch einen langen Aufenthalt und eine erfolgreiche Bildungsgeschichte in Deutschland nachgewiesen ist und sie darüber hinaus in einem anderen Test der globalen Sprachfertigkeit vergleichbar mit Muttersprachlern abschneiden, liegt die Leistung von *Near-Natives* in der Aufgabe Satzwiedergabe unter der von Muttersprachlern.

Allerdings wäre auch eine alternative Erklärung dieses ersten Ergebnisses möglich: Die Satzwiedergabe könnte ein sensiblerer Indikator für Unterschiede in der Sprachkompetenz sein als der eingesetzte C-Test. Alle Leistungen in der Satzwiedergabe, die innerhalb einer bestimmten Spanne liegen, würden dann in einer weit geringeren Spanne des C-Tests liegen oder sogar das gleiche Ergebnis im C-Test hervorbringen. Der C-Test wäre damit nicht ausreichend diskriminierend, um die Unterschiede abzubilden, die die Satzwiedergabe zeigt. Damit wäre die Satzwiedergabe der bessere Test. In diesem Fall sollten aber im Umkehrschluss alle Teilnehmer, die einen vergleichbaren Wert in der Satzwiedergabe zeigen, auch den gleichen C-Test-Wert erreichen.

Um zu testen, ob dies zutrifft, wurde eine Parallelisierung anhand der Wiedergabeleistung durchgeführt. Dazu wurde eine Gruppe jugendlicher Muttersprachler getestet, die den Muttersprachlern aus Studie 2 demographisch möglichst ähnlich sind, von denen aber aufgrund ihres Alters schlechtere Wiedergabeleistungen erwartet wurden. Dieser Altersunterschied musste hingenommen werden, da entsprechend niedrige Leistungen bei erwachsenen Muttersprachlern mit vergleichbarem Bildungshintergrund nicht zu erwarten sind. Die Schüler besuchten jedoch zum Testzeitpunkt ein Gymnasium, mit dessen erfolgreichem Abschluss sie eine Hochschulzugangsberechtigung erwerben. Demnach sollten sie in wichtigen Faktoren (wie z.B. der allg. Intelligenz im Vergleich zu ihrer Alterskohorte) mit den Studenten vergleichbar sein. Im Folgenden soll außerdem dargelegt werden, warum der Altersunterschied nicht problematisch für den Zweck der Studie ist. Beide Aufgaben (C-Test und Satzwiedergabe) sind mit dem Ziel konstruiert, globale Sprachfertigkeit zu messen und erfüllen diese Anforderung nachweislich für Muttersprachler (siehe oben). Im Alter der getesteten jugendlichen Muttersprachler (13-14 Jahre) sind die Sprachfertigkeiten voll entwickelt, mit einem Nachteil für das Lexikon, das sich über die Lebensspanne stetig weiter entwickelt. Es gibt keinen Grund für die Annahme, dass eine der beiden Aufgaben das Lexikon in höherem Maße beansprucht als die andere, also sollte sich dieser Unterschied zwischen jugendlichen und erwachsenen Muttersprachlern in beiden Tests *gleichermaßen* niederschlagen. Weiterhin ist für die

Dissoziation der beiden Maße C-Test und Satzwiedergabe einzig entscheidend, ob Teilnehmer, die gleichauf in ihrer Leistung in der Satzwiedergabe liegen, auch gleich gut im C-Test abschneiden. Die Logik der Dissoziation der beiden Tests in beide Richtungen (gleiche Leistungen im C-Test & Unterschiede in der Satzwiedergabe; gleiche Leistungen in der Satzwiedergabe & Unterschiede im C-Test) ist nicht davon abhängig, auf welcher Dimension sich die Teilnehmergruppen unterscheiden.

Insgesamt deutet die Analyse der unterschiedlich parallelisierten Daten darauf hin, dass die Aufgabe „Sätze wiedergeben“ die sprachlichen Fähigkeiten hochkompetenter L2-Sprecher nicht valide abbilden kann, zumindest nicht, wenn diese Leistung mit der von Muttersprachlern verglichen wird.

7.4 *Gemeinsame Diskussion der Studien 1 und 2*

In den Studien 1 und 2 wurden jeweils bilinguale mit monolingualen Personen verglichen. Es konnte gezeigt werden, dass sprachlich kompetente Kinder mit DaZ und erwachsene *Near-Natives* im Vergleich zu Muttersprachlern schlechter in der Satzwiedergabe abschneiden, als anhand ihrer Leistung in alternativen Tests der Sprachkompetenz zu erwarten war. Dieser Vergleich ist jedoch problematisch, da die Sprachverarbeitung bilingualer Sprecher grundsätzlich, also auch für ihre L1, unterschiedlich von der monolingualer Sprecher sein könnte. Daher soll in Studie 3, die den ersten empirischen Teil dieser Arbeit abschließt, gezeigt werden, dass sich Unterschiede in der Satzwiedergabe in einer L2 auch in einer gänzlich bilingualen Gruppe mit unterschiedlicher Erst- und Zweitsprache aufzeigen lässt. Dies ist nötig, um zu zeigen, dass die Verarbeitungsunterschiede auf die Verarbeitung der L2 im Vergleich zur L1 zurück zu führen sind, nicht auf die Bilingualität bzw. Monolingualität der Sprecher an sich.

Weiter oben bereits referierte Studien zeigen eine weniger automatisierte Sprachverarbeitung für eine L2 auf, jedoch wurden auch in diesen Studien meist monolinguale Sprecher mit bilingualen verglichen. Für einige Aufgabentypen konnten auch Vorteile für bilinguale Sprecher im Vergleich zu monolingualen aufgezeigt werden. Bialystok (2007; s. auch Bialystok, Craik, Klein & Viswanathan, 2004; Luo, Luk & Bialystok, 2010) berichtet von Vorteilen für bilinguale Sprecher im Vergleich zu monolingualen in einigen Funktionen der kontrollierten Verarbeitung. Diese Vorteile

könnten die Nachteile durch die geringere Automatisierung abschwächen oder aufheben. Im Folgenden soll kurz zusammengefasst werden, warum diese Vorteile jedoch nicht bei der Satz wiedergabe zu erwarten sind.

Vorteile in der kontrollierten Verarbeitung für bilinguale Probanden finden sich zumeist dann, wenn irrelevante Information unterdrückt werden muss, ein Konflikt zwischen unterschiedlichen Impulsen aufgelöst werden muss oder zwischen verschiedenen Systemen von Regeln gewechselt werden muss (vgl. Bialystok, 2007). Dies trifft nicht auf die Wiedergabe von Sätzen zu, bei der die Information lediglich abrufbar gehalten wird. Auch konnte der Effekt fast ausschließlich in der Verarbeitung nonverbalen Materials gezeigt werden. Eine Ausnahme davon bildet eine Studie von Luo, Luk und Bialystok (2010), in der ein Vorteil in der Flüssigkeit der Produktion von Wörtern für bilinguale Sprecher gefunden wurde. Eine Gruppe einsprachiger Teilnehmer wurde anhand zweier Aufgabenstellungen mit einer Gruppe bilingualer Teilnehmer verglichen. Die Teilnehmer wurden zum einen angewiesen, Wörter zu produzieren, die zu einer bestimmten semantischen Kategorie gehörten und zum anderen Wörter, die mit einem bestimmten Buchstaben begannen. Es zeigte sich, dass beide Gruppen gleich abschnitten, wenn Wörter aus einer semantischen Kategorie zu produzieren waren, bilinguale Sprecher jedoch im Vorteil waren, wenn Wörter mit dem gleichen Anfangsbuchstaben produziert werden sollten (vgl. Luo et al., 2010). Luo et al. interpretieren dies als besser ausgeprägte Fähigkeit der Inhibition bei bilingualen Sprechern. Inhibition sei beim Produzieren von Wörtern mit dem gleichen Anfangsbuchstaben nötig, da das Produzieren semantisch verwandter Begriffe, das näher an der Architektur des Sprachsystems sei, während der Produktion phonologisch verwandter Begriffe unterdrückt werden muss. Die besser eingeübte Inhibition durch die ständige Unterdrückung der gerade nicht relevanten Sprache bei den bilingualen Teilnehmern sei für deren Vorteil in dieser Aufgabe verantwortlich. Obwohl in diesem Fall ein Vorteil bilingualer Sprecher mit einer verbalen Aufgabenstellung gezeigt werden konnte, unterscheiden sich die Ansprüche dieser Aufgabe an die kognitiven Prozesse stark. Diese Wortproduktionsaufgaben beinhalten keinerlei Behaltensleistung, welche bei der Satz wiedergabe jedoch ein kritischer Faktor ist. Gleichzeitig kommt es bei der Satz wiedergabe nur in beschränktem Maße darauf an, irrelevante Information zu unterdrücken. Inhibition spielt jedoch in der zweiten Wortproduktionsaufgabe, in der ein Vorteil für bilinguale Teilnehmer gefunden wurde, eine zentrale Rolle.

Dennoch soll in Studie 3 ausgeschlossen werden, dass die Verarbeitungsunterschiede zwischen Muttersprachlern und Zweitsprachlern von grundlegenden Unterschieden zwischen monolingualen und bilingualen Sprechern beeinflusst werden. Dazu wird die Leistung in der L1 und L2 einer vollständig bilingualen Teilnehmergruppe verglichen.

Daneben soll gezeigt werden, dass der Effekt grundsätzlich auch auf andere Sprachen (hier das Französische) übertragen werden kann. In Studie 3 soll daher auch die Leistung der Teilnehmer in ihrer L1 bzw. L2 Französisch erfasst werden. So wird geprüft, ob der in Studie 2 aufgedeckte Effekt auch für das Französische gezeigt werden kann. Die Teilnehmer in Studie 3 bilden außerdem eine sehr homogene Gruppe, was potentielle Störvariablen wie den sozialen Hintergrund und das Alter, sowie das Erwerbsalter und die Spracherwerbsgeschichte ihrer L2 angeht.

8 Studie 3 - Satzbehalten bilingualer Schüler in Deutsch und Französisch

In dieser Studie werden in einem 2x2-Design die Leistungen in der Satzwiedergabe und dem C-Test in der L1 und L2 einer durchweg bilingualen Teilnehmergruppe verglichen. Analog zu Studie 1 wird der Faktor TEST (C-Test vs. Satzwiedergabe) mit dem Faktor L1 (Muttersprache - Zweitsprache) gekreuzt. Die Teilnehmer sprechen dabei entweder Deutsch als L1 und Französisch als L2 oder Französisch als L1 und Deutsch als L2 und können daher in beiden Sprachen getestet werden, die Studie verwendet also französisches und deutsches Material. Die Muttersprachler in der französischen Version sind demnach die Zweitsprachler in der deutschen Version und umgekehrt. Damit soll getestet werden, ob es sich bei dem Effekt aus Studie 2 um einen generellen Nachteil für bilingualen Sprecher handelt oder das schlechtere Abschneiden wie erwartet auch bei bilingualen Sprechern auf die Zweitsprache beschränkt ist. Es wird erwartet, dass Letzteres der Fall ist und sich auch in dieser Stichprobe ein Unterschied in der Verarbeitung der L1 im Vergleich zur L2 finden lässt, der darüber hinaus in beiden Sprachen, Deutsch und Französisch, präsent ist. Es wird erwartet, dass sich für beide Sprachen die jeweiligen L1- und L2-Sprecher in der Satzwiedergabe stärker unterscheiden als im C-Test, wobei in der L2 eine allgemein schlechtere Leistung erwartet wird als in der L1. Dies soll sich statistisch in einem Haupteffekt für L1 und einer Interaktion zwischen den Variablen L1 und TEST zeigen.

8.1 Methode

Stichprobe

Es nahmen 48 bilingualen Schüler des DFG/LFA (Deutsch-Französisches Gymnasium/ Lycée Franco-Allemand) in Saarbrücken teil. Das Gymnasium liegt in direkter Nähe zu Frankreich und wird sowohl von Deutsch- als auch Französischmuttersprachlern besucht. Die Schüler erhalten mit dem Beginn des Besuchs der Schule in Klassenstufe 5 verstärkten Unterricht in ihrer jeweiligen L2, der ab Klassenstufe 8 in einen integrierten Unterricht (Schüler mit L1 Deutsch und L1 Französisch in gemeinsamen Klassenverbänden) und einem Gleichgewicht der beiden Sprachen im Unterricht mündet. Im schulischen Leben jenseits des Unterrichts wird von der Schulleitung und dem deutsch-französischen Kollegium ebenfalls auf eine ausgewogene Verwendung der Sprachen Deutsch und Französisch geachtet. Da die

Schüler zwar im Grenzgebiet zwischen Frankreich und Deutschland leben, aber zumeist keine Immersion in ihrer jeweiligen L2 außerhalb des schulischen Lebens erfahren, ist dennoch von allgemeinen Leistungsunterschieden zwischen ihrer Erst- und Zweitsprache auszugehen, auch wenn sie hochkompetente Sprecher ihrer jeweiligen L2 sind.

Zum Zeitpunkt der Erhebung standen die teilnehmenden Schüler am Ende des 11. Schuljahres (ein Jahr vor dem Abitur/Baccalauréat) und waren 17 Jahre alt. Alle Schüler besuchten das DFG/LFA seit der Klassenstufe 5 (ab dem Alter von 10 Jahren). Von den 48 teilnehmenden Schülern sprechen 29 Deutsch als L1 (Gruppe D, 20 weiblich), neun sprechen Französisch als L1 (Gruppe F, sechs weiblich), sieben sind simultan mit Deutsch und Französisch aufgewachsen (Gruppe S, vier weiblich) und drei sprechen weder Deutsch noch Französisch als L1 (Gruppe N, zwei weiblich). Da alle Hypothesen und theoretischen Annahmen dieser Arbeit sich auf den sukzessiven Zweitspracherwerb beziehen, wurde die Gruppe S (simultan bilingual) von der weiteren Analyse ausgeschlossen. Gruppe N wurde ebenfalls von der Analyse ausgeschlossen, da weder Deutsch noch Französisch L1 dieser Teilnehmer war. Die Erstsprachen dieser Schüler waren Türkisch, Persisch und Englisch; Deutsch erwarben sie ab dem Alter von 3-5 Jahren, Französisch ab dem Alter von 10 Jahren¹⁹. Im Vorfeld hatten alle Schüler einen Fragebogen zu ihrer Spracherwerbsgeschichte ausgefüllt und eine unterschriebene Einverständniserklärung ihrer Eltern vorgelegt, sowie ihr eigenes schriftliches Einverständnis zur Teilnahme gegeben. Die Schüler erhielten eine Aufwandsentschädigung für ihre Teilnahme.

Material

In dieser Studie werden 20 der deutschen Sätze (in ihrer schriftlichen und auditiven Form) und der deutsche C-Test aus Studie 2 verwendet. Daneben wurden 20 Sätze aus der französischen Version des Lesespannentests von Daneman und Carpenter (Desmette, Hupet, Schelstraete & van der Linden, 1995) ausgewählt und von einer französischen Muttersprachlerin um einzelne Wörter ergänzt, so dass die Sätze nur geringfügig verändert werden und ihre Wortanzahl der der deutschen Sätze entsprach. Ein französischer C-Test, der für den Einsatz mit Muttersprachlern entwickelt worden war (Grotjahn, 1992), wurde in gekürzter Form eingesetzt. Er bestand aus fünf Kurztexten, von denen vier 20 Lücken enthalten und einer 19 Lücken (s. Anhang A4). Somit waren die maximal erreichbaren

¹⁹ Auf die Möglichkeit, diese Gruppe in beiden Studien als Zweitsprachler zu behandeln wurde verzichtet, um die Versuchsgruppe hinsichtlich ihrer Sprachgeschichte möglichst homogen zu halten.

Werte im deutschen und französischen C-Test gleich. Eine französische Muttersprachlerin erstellte Audioaufnahmen der 20 Sätze (s. Anhang A3) und mehrerer Übungssätze.

Tabelle 5: Balance der Sprachen und Aufgaben im Design von Studie 3.

Abfolge 1	Abfolge 2	Abfolge 3	Abfolge 4
C-Test Deutsch	C-Test Franz.	Deutsch-aud.	Franz.-aud.
C-Test Franz.	C-Test Deutsch	Franz.-aud.	Deutsch-aud.
Deutsch-aud.	Franz.-aud.	C-Test Deutsch	C-Test Franz.
Franz.-aud.	Deutsch-aud.	C-Test Franz.	C-Test Deutsch

Ablauf

Die Schüler wurden an Laptops in einem ruhigen Klassenzimmer am DFG/LFA getestet. Das Material wurde wie in Studie 2 auditiv (über Kopfhörer) dargeboten. Die Satzwiedergabe erfolgte schriftlich auf vorbereiteten Antwortbögen. Die Sätze wurden in 2 Blöcken mit jeweils 10 Sätzen getrennt nach Sprachen (Deutsch – Französisch) präsentiert. Vor jedem Block bearbeiteten die Teilnehmer zwei Übungsdurchgänge, um sich mit dem Ablauf vertraut zu machen. Die Verwendung der Sätze war über die Probanden hinweg ausbalanciert. Für die Hälfte der Teilnehmer war Französisch die zuerst dargebotene Sprache. Innerhalb der Blöcke wurden die Sätze randomisiert dargeboten. Die Hälfte der Teilnehmer bearbeitete die C-Tests vor der Satzwiedergabe, die Hälfte die Satzwiedergabe vor den C-Tests; die Hälfte der Teilnehmer bearbeitete den französischen C-Test vor dem deutschen C-Test. (Einen Überblick über den Ablauf gibt die Tabelle 5.) Insgesamt dauerte eine Sitzung ca. 45 Minuten.

Auswertung

Die Auswertung der C-Tests und der Satzwiedergabe erfolgte analog zur Auswertung in Studie 2.

8.2 Ergebnisse

Schüler mit L1 Deutsch schnitten im deutschen C-Test besser ab ($M = 93,8$, $SD = 3,22$), als Schüler mit L1 Französisch ($M = 76,67$, $SD = 11,61$). Umgekehrt war im französischen C-Test die Leistung der Schüler mit L1 Französisch besser ($M = 85,00$, $SD = 7,31$) als die der Schüler mit L1 Deutsch ($M = 65,14$, $SD = 11,80$). Die statistische Analyse erfolgte analog zu Studie 1 mittels einer ANOVA mit Messwiederholung. Im Folgenden werden die Ergebnisse für beide Sprachen getrennt berichtet.

Tabelle 6: Leistungen in der deutschen Satz wiedergabe und dem deutschen C-Test für Schüler mit L1 Französisch und L1 Deutsch.

	C-Test (deutsch)	Satzwiedergabe (deutsch)
L1 Französisch	Mean = 76,67 SD = 11,608	Mean = 57,19 SD = 18,210
L1 Deutsch	Mean = 93,8 SD = 3,22	Mean = 83,76 SD = 7,58

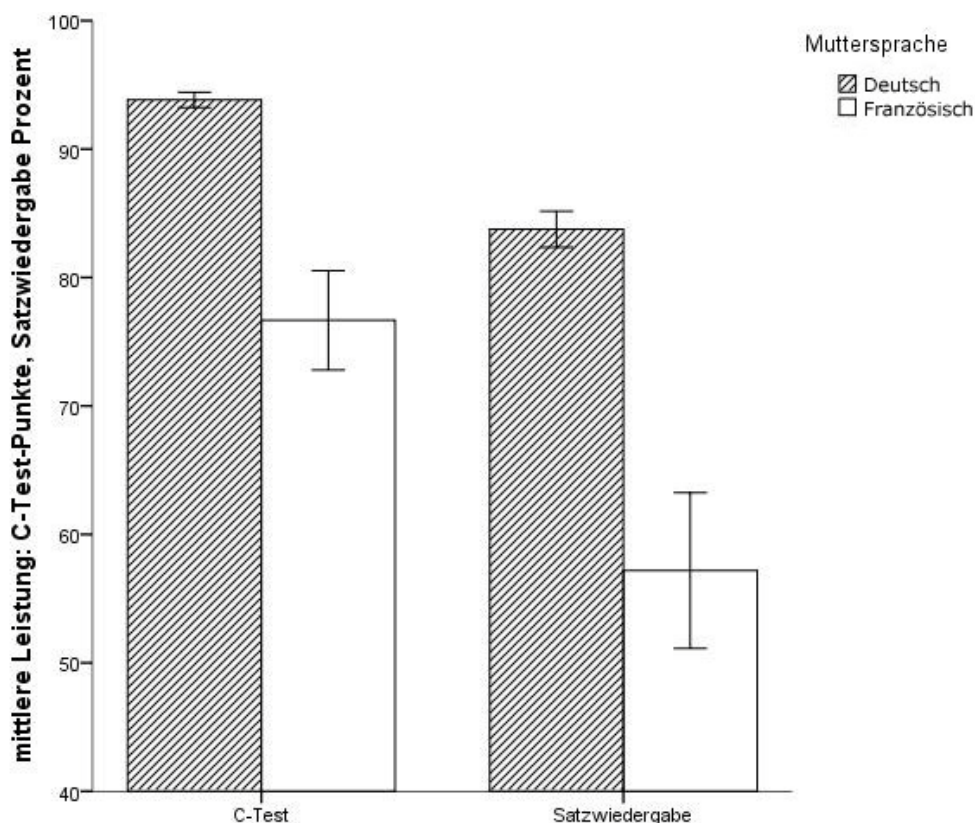


Abbildung 9: Vergleich der Leistung bilingualer Schüler mit L1 Deutsch bzw. L1 Französisch im deutschen C-Test und der Wiedergabe deutscher Sätze (als Prozent der korrekt wiedergegebenen Wörter).

Tabelle 7: Leistungen in der deutschen Satz wiedergabe und dem deutschen C-Test für Schüler mit L1 Französisch und L1 Deutsch.

	C-Test (franz.)	Satzwiedergabe (franz.)
L1 Französisch	Mean = 85 SD = 7,31	Mean = 84,42 SD = 10,95
L1 Deutsch	Mean = 65,14 SD = 11,8	Mean = 50,83 SD = 18,22

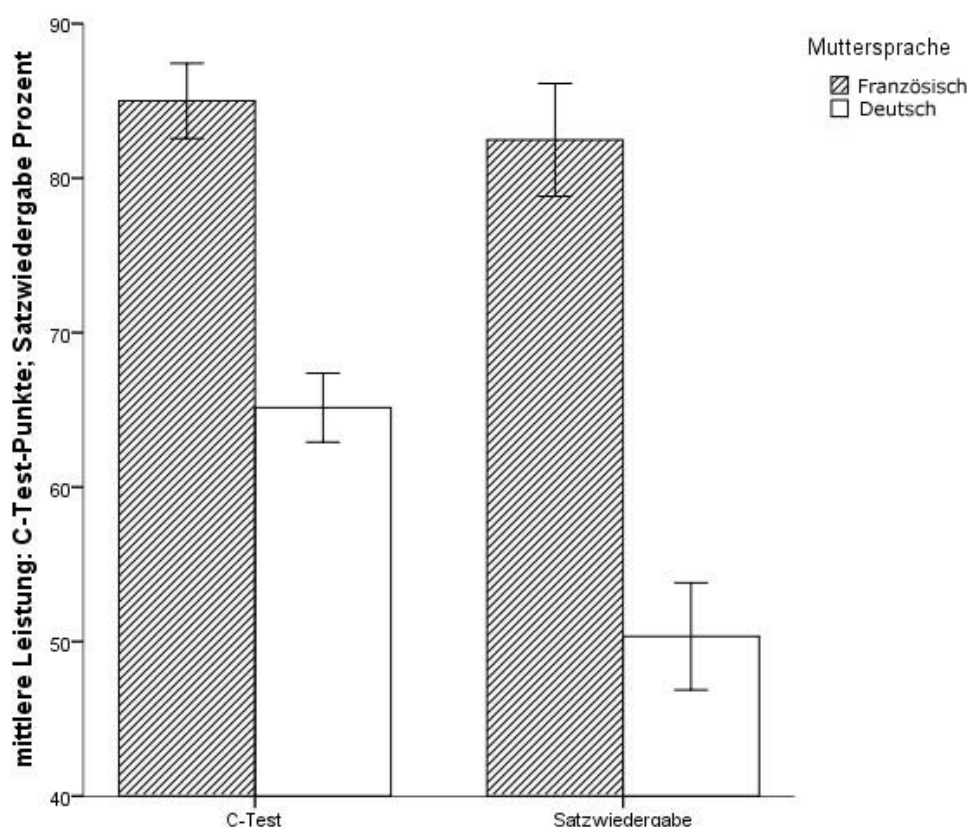


Abbildung 10: Vergleich der Leistung bilingualer Schüler mit Französisch als L1 bzw. L1 Deutsch im französischen C-Test und der Wiedergabe französischer Sätze (als Prozent der korrekt wiedergegebenen Wörter).

Deutsch

Im deutschen C-Test schnitten Schüler mit L1 Französisch schlechter ab als Schüler mit L1 Deutsch. Schüler mit L1 Deutsch schnitten besser in der Satz wiedergabe ab als Schüler mit L1 Französisch (s. Tabelle 6).

Zur Analyse wurde eine ANOVA mit Messwiederholung mit den Faktoren L1 (L1 Deutsch – L2 Deutsch) und TEST (deutscher C-Test – deutsche Satz wiedergabe) durchgeführt. Schüler mit L1 Deutsch (n = 29) und Schüler mit L1 Französisch (n = 9)

unterschieden sich in beiden Aufgabentypen, was sich in einem Haupteffekt für L1 zeigte ($F(1, 36) = 53,20; p < .001$). Daneben zeigte sich ein Haupteffekt für TEST mit $F(1, 36) = 91,17; p < .001$. Zwischen den Faktoren L1 und TEST zeigte sich eine Interaktion ($F(1, 39) = 9,24; p < .005$). Geplante Einzelvergleiche zeigten, dass sich die beiden Gruppen in ihrer C-Testleistung ($F(1,36) = 53,22; p < .001$) und in der Satzwiedergabe voneinander unterschieden ($F(1,36) = 40,97; p < .001$), was sich im Haupteffekt für L1 bereits andeutete.

Französisch

Schüler mit L1 Französisch schnitten im französischen C-Test besser ab als Schüler mit L1 Deutsch. In der Wiedergabe der französischen Sätze lag die Leistung der Schüler mit L1 Französisch über der Leistung der Schüler mit L1 Deutsch (vgl. Tabelle 7).

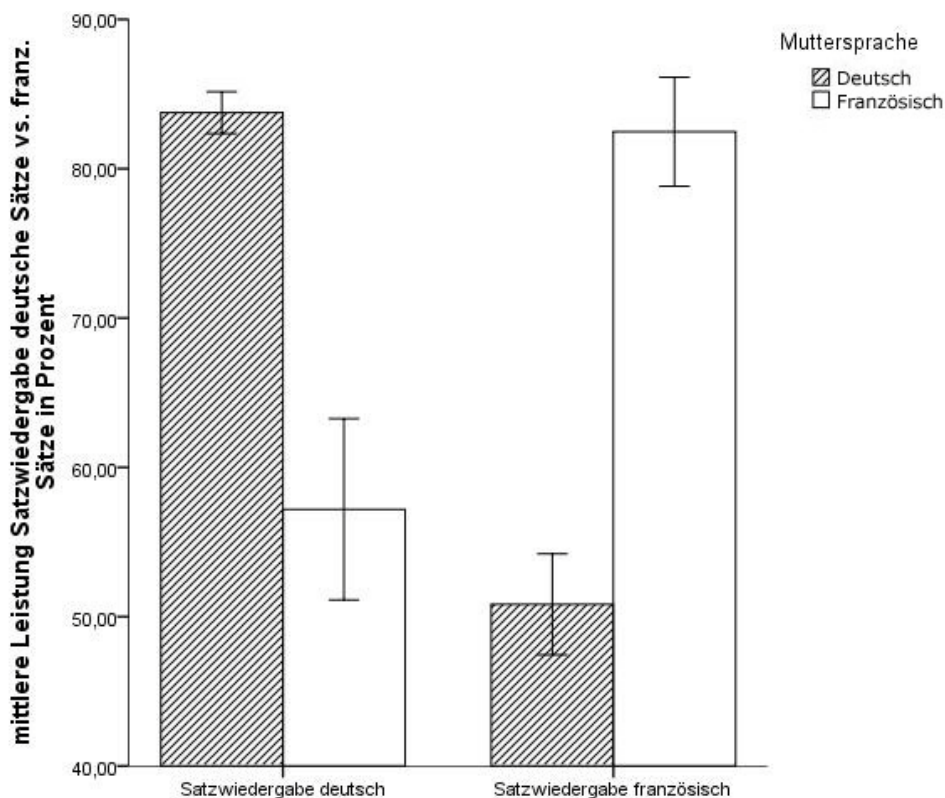


Abbildung 11: Vergleich der Leistung bilingualer Schüler mit Französisch als L1 bzw. Deutsch als L1 in der Wiedergabe französischer und deutscher Sätze (als Prozent der korrekt wiedergegebenen Wörter).

Eine ANOVA mit Messwiederholung wurde mit den Faktoren L1 (Deutsch – Französisch) und TEST (französischer C-Test – französische Satzwiedergabe) durchgeführt. Dabei zeigte sich, dass insgesamt die Schüler mit L1 Französisch ($n = 9$) in

ihrer Leistung den Schülern mit L1 Deutsch ($n = 28$)²⁰ signifikant überlegen waren, $F(1, 35) = 27,9; p < .001$. Weiterhin unterschieden sich die Leistungen in den beiden Aufgabentypen über die Sprachgruppen hinweg, $F(1, 35) = 13,25; p = .001$. Die Interaktion der beiden Faktoren war signifikant ($F(1, 35) = 6,66; p < .02$). Geplante Einzelvergleiche zeigten, dass die Gruppen sich in der Satz wiedergabe ($F(1,35) = 24.47; p < .001$) und im C-Test unterschieden, $F(1,35) = 22.44; p < .001$.

Deutsche vs. Französische Satz wiedergabe

Wie bereits berichtet wurde, war die jeweilige L1-Gruppe der entsprechenden L2-Gruppe in der Satz wiedergabe in beiden Sprachen überlegen. Um diesen Unterschied noch einmal systematischer zu betrachten, wurde eine 2x2-ANOVA mit Messwiederholung mit den Faktoren TESTSPRACHE (deutsch vs. französische Satz wiedergabe) und L1 (Deutsch vs. Französisch) gerechnet. Dabei zeigten sich keine Haupteffekte für TESTSPRACHE ($F(1,36) = 1,2$) und L1 ($F < 1$). Die Interaktion der beiden Faktoren war jedoch signifikant, $F(1,36) = 71,9; p < .001$. In geplanten Einzelvergleichen zeigten sich Unterschiede zwischen den Sprachgruppen in der deutschen Satz wiedergabe, $F(1,36) = 40,97; p < .001$, und der französischen Satz wiedergabe, $F(1,36) = 24,14; p < .001$. Die Leistungen innerhalb der Sprachgruppen in den beiden Maßen der Satz wiedergabe unterschieden sich ebenfalls signifikant. Die Gruppe mit L1 Französisch schnitt in der französischen Satz wiedergabe besser ab als die Gruppe L1 Deutsch, $F(1,36) = 17,79; p < .001$. Die Gruppe mit L1 Deutsch schnitt in der deutschen Satz wiedergabe besser ab als die Gruppe mit L1 Französisch, $F(1,36) = 97,13; p < .001$.

8.3 *Diskussion*

Auch im Vergleich von Deutsch-Französisch und Französisch-Deutsch bilingualen Schülern konnte ein Nachteil für die Leistung in der L2 in der Satz wiedergabe gezeigt werden. Gleichzeitig konnte gezeigt werden, dass sich dieser Nachteil nicht nur in der deutschen Sprache finden lässt, sondern auf das Französische übertragen lässt. Durch den Vergleich innerhalb einer Gruppe von bilingualen Sprechern wird (zusammen mit den Ergebnissen aus den Studien 1 und 2) sichergestellt, dass es sich bei dem Nachteil in der Satz wiedergabe um einen spezifischen Nachteil für die L2 handelt.

²⁰ Eine Probandin hatte den französischen C-Test nicht ausgefüllt und konnte in dieser Analyse nicht berücksichtigt werden.

Das schlechtere Abschneiden in der Satzwiedergabe im Vergleich zur Leistung im C-Test für Sprecher einer L2 im Vergleich zu Muttersprachlern kann aufgrund dieser Ergebnisse und der Ergebnisse der beiden vorangegangenen Studien nicht auf einen Nachteil für bilinguale Sprecher reduziert werden. Der Effekt geht vielmehr auf die Verarbeitungsunterschiede zwischen einer L1 und einer L2 zurück. Gleiches gilt für den Einfluss eines möglichen Vorteils bilingualer Probanden durch höhere Leistungsfähigkeit bestimmter Funktionen der zentralen Exekutive (vgl. Bialystok, 2007). Weiter oben wurden bereits Unterschiede zwischen den Anforderungen der Satzwiedergabe und den Anforderungen der Aufgabentypen aufgezeigt, für die sich ein Vorteil für bilinguale gegenüber monolingualen Sprechern zeigen lässt (auch: Bialystok, 1997; Luo et al. 2010). Die Ergebnisse aus Studie 3 mit einer gänzlich bilingualen Stichprobe sind Evidenz dafür, dass es sich bei dem gefundenen Effekt um Verarbeitungsunterschiede zwischen einer L1 und einer L2 handelt und nicht um Verarbeitungsunterschiede, die allgemein zwischen bilingualen und monolingualen Sprechern bestehen. Entsprechend der Testsprache (Deutsch - Französisch), die gerade betrachtet wird, sind die selben Teilnehmer einmal in der schlechter und einmal in der besser abscheidenden Gruppen, je nachdem, ob ihre L1 oder L2 betrachtet wird.

Nicht optimal ist die sehr ungleiche Anzahl von Versuchsteilnehmern mit L1 Deutsch (29 Probanden) und L1 Französisch (9 Probanden). Trotz der Ansprache von Schülern und Eltern in beiden Sprachen war die Teilnahmebereitschaft der deutschen Schüler weitaus größer als die der französischen Schüler. Über die Gründe könnte ich nur spekulieren. Die unterschiedliche Gruppengröße ist eine Schwäche der Studie, die leider durch das nicht vorhersehbare ungleiche Verhalten der Schüler entstanden ist. Da die Studie auch für die Schule mit Aufwand verbunden war und in den schulischen Ablauf integriert werden musste, war eine Nacherhebung und ein verstärktes Anwerben französischer Schüler als Teilnehmer leider nicht umzusetzen.

8.4 *Gemeinsame Diskussion Studien 1-3*

In drei Studien konnte gezeigt werden, dass die Leistung im Satzbehalten sich unterschiedlich von der Leistung in anderen Sprachtests unterscheidet, je nachdem ob in einer L1 oder einer L2 gemessen wird. Diese Grundannahme kann durch die drei Studien bestätigt werden. Selbst *Near-Natives*, die ihre L2 auf höchstem Niveau beherrschen und auf muttersprachlichem Niveau im C-Test abschnitten, zeigten in der Satzwiedergabe weit

schlechtere Leistungen als Muttersprachler. Die erste Studie belegte dies für die Zielgruppe der derzeit in Deutschland durchgeführten Sprachstandserhebungen, Kinder mit DaM und DaZ, die ungefähr zwei Jahre vor ihrer Einschulung stehen. In der zweiten Studie zeigten sich selbst zwischen *Near-Natives* und Muttersprachlern, die im C-Test gleich abschnitten, Unterschiede in der Satzwiedergabe. In einer dritten Studie mit einer durchweg bilingualen Stichprobe konnte darüber hinaus gezeigt werden, dass sich der aufgabenspezifische Nachteil in der Satzwiedergabe nicht durch die Zweisprachigkeit per se oder individuelle Unterschiede im Satzbehalten erklären lässt, sondern auf das Testen in einer L2 beschränkt ist. Daneben konnten in der dritten Studie die Ergebnisse auch auf die französische Sprache übertragen werden.

Bei jeder einzelnen der Studien bestehen jedoch Schwierigkeiten bei der Interpretation der Ergebnisse. In Studie 1 ist die Zugehörigkeit zu den beiden Sprachgruppen (DaM und DaZ) mit dem Alter der Kinder konfundiert. Genauere Untersuchungen des Zusammenhangs zwischen Alter und Satzwiedergabeleistung durch eine multiple Regression und eine separate Analyse der Kinder mit DaZ deuten jedoch darauf hin, dass nicht das Alter der Kinder sondern die Zugehörigkeit zur Sprachgruppe DaM bzw. DaZ den Effekt besser erklären. Auch ist die Gruppengröße der Kinder mit DaZ und DaM unterschiedlich. Dies ergab sich, ähnlich wie auch in Studie 3, durch die Notwendigkeit der Kooperation mit Kindertagesstätten, dem Einverständnis der Eltern und der Kooperationsbereitschaft der Kinder selbst. All diese Voraussetzungen für die Erhebung der Daten macht es schwer möglich, die Versuchsgruppen gemäß ihres Alters konstant zu halten.

In Studie 2 ist der Vergleich zwischen erwachsenen *Near-Natives* und jüngeren Muttersprachlern problematisch. Er ergibt sich aus der Notwendigkeit, eine Vergleichsgruppe heranzuziehen, die sich hinsichtlich der Wiedergabeleistung mit den *Near-Natives* vergleichen lässt. Bei Muttersprachlern mit ähnlichem Alter und Bildungsniveau kann dieses Leistungsniveau nicht gefunden werden.

Trotz der bestehenden Schwächen der einzelnen Studien, die sich hauptsächlich aus den Bedürfnissen der Institutionen herleiten, mit denen für die Studien kooperiert wurde, ergibt sich aus ihnen insgesamt konvergierende Evidenz für die Annahme, dass sich Sprecher einer L2 in ihrer Leistung in der Satzwiedergabe stärker von Sprechern einer L1 unterscheiden als in anderen Aufgabentypen, dass also *Near-Natives* im Satzbehalten schlechter abschneiden im Vergleich zu Muttersprachlern als in alternativen Maßen der

Sprachkompetenz. Dies lässt sich theoretisch durch die besonderen Anforderungen der Satzwiedergabe an die Aufmerksamkeit erklären.

Während die drei Studien im ersten Teil dieser Arbeit darauf abzielten aufzuzeigen, dass auch hochkompetente Sprecher einer L2 schlechtere Leistung im wörtlichen Behalten von Sätzen zeigen als aufgrund ihrer Leistung in anderen Tests zu erwarten ist, soll nun in den folgenden Studien geprüft werden, ob es möglich ist, eine Alternative zur wörtlichen Satzwiedergabe in Sprachtests einzusetzen. Es wäre eine alternative Aufgabe wünschenswert, die die Vorteile der Satzwiedergabe beibehält, jedoch den Nachteil der hohen Belastung des Arbeitsgedächtnisses nicht aufweist. Eine solche Aufgabe könnte die inhaltliche Satzwiedergabe sein. Voraussetzung dafür wäre, dass das inhaltliche Behalten eine niedrigere Belastung des Arbeitsgedächtnisses mit sich bringt als das wörtliche Behalten eines Satzes. Im ersten Schritt (Studie 4) wird daher die Vorannahme einer geringeren Belastung des Arbeitsgedächtnisses beim Lesen zur inhaltlichen Wiedergabe im Vergleich zum Lesen zur wörtlichen Wiedergabe getestet. Dies geschieht zunächst für die muttersprachliche Verarbeitung. Im zweiten Schritt (Studie 5) wird die Leistung von Muttersprachlern und *Near-Natives* in der wörtlichen und inhaltlichen Satzwiedergabe, sowie dem C-Test als alternativem Maß der Sprachkompetenz geprüft. Diese beiden Studien bilden den zweiten Teil dieser Arbeit.

9 Studie 4 - Wörtliches und inhaltliches Satzbehalten in der L1

Im ersten Teil dieser Arbeit wurde argumentiert und empirisch untermauert, dass die wörtliche Satz wiedergabe als Maß der Sprachkompetenz für die Messung in einer L2 nicht geeignet ist, da die Sprachfertigkeit kompetenter Sprecher mit diesem Maß vermutlich systematisch unterschätzt wird. Dies ist auf die aufgabenspezifische hohe Belastung der Aufmerksamkeit zurückzuführen. Diese hohe Belastung führt dazu, dass Unterschiede im Automatisierungsgrad zwischen einer L1 und L2 die Leistung in der Satz wiedergabe beeinflussen. Dieser niedrigere Automatisierungsgrad zeigt sich bei hochkompetenten Sprechern einer L2 ausschließlich bei einer ungewöhnlich hohen Belastung der Aufmerksamkeit. Er sollte nicht als Unterschied in der allgemeinen Sprachkompetenz interpretiert werden, insbesondere wenn in alternativen Tests der Sprachfähigkeit und im Alltag ein muttersprachliches Niveau erreicht wurde.

Ziel des zweiten Teils dieser Arbeit ist es, eine Aufgabe zu identifizieren, mit der die Sprachfertigkeit von Sprechern einer L2 erfasst werden kann und die die wörtliche Satz wiedergabe als Aufgabe ersetzen könnte. Ideal wäre, wenn diese Aufgabe die Vorteile der Satz wiedergabe beibehält und eine weniger hohe Belastung der Aufmerksamkeit mit sich bringt. Diese Anforderungen könnten durch die inhaltliche (im Kontrast zur wörtlichen) Satz wiedergabe erfüllt sein. Bei der inhaltlichen Satz wiedergabe beschränkt sich die geforderte Behaltensleistung auf den Inhalt, also den propositionalen Gehalt, des Satzes. Die Satzoberfläche muss dabei nicht in einem Zustand der Aktivierung gehalten werden, die Aufmerksamkeit wird vermutlich geringer belastet.

Aaronson und Scarborough (1976) fanden in einer Lesestudie Unterschiede in der Lesegeschwindigkeit für Sätze, je nachdem ob diese (nur) verstanden oder (auch) behalten werden sollten. Die Lesegeschwindigkeit wurde mittels eines *self paced reading* Aufbaus gemessen. Die Sätze wurden Wort für Wort am Bildschirm präsentiert, das nächste Wort konnte durch das Drücken einer Taste durch den Versuchsteilnehmer gesteuert werden. Die Abstände zwischen dem wiederholten Drücken der Taste (also die Zeit, für die ein Wort am Bildschirm angezeigt wurde) wurde als Lesezeit für dieses Wort interpretiert. Die Lesezeiten der einzelnen Wörter ergeben gemittelt oder als Verlauf betrachtet die Lesegeschwindigkeit des Satzes. Versuchsteilnehmer, die instruiert wurden, sich Sätze wörtlich zu merken, lasen diese langsamer als Versuchsteilnehmer, die die Sätze lediglich verstehen sollten. Auch unterschied sich die Lesegeschwindigkeit über den Satz hinweg:

Mit der Instruktion „Behalten“ wurde die Lesegeschwindigkeit über den Satz hinweg geringer, mit der Instruktion „Verstehen“ hingegen wurde sie höher. Aaronson und Scarborough (1976) erklärten Ersteres mit der steigenden Beanspruchung des Arbeitsgedächtnisses: Über die Verarbeitungsdauer des Satzes hinweg muss immer mehr Information behalten werden. Diese Behaltensleistung beansprucht Ressourcen, die nicht mehr für die Verarbeitung zur Verfügung stehen. Dadurch wird die Verarbeitung des Satzes und damit die Lesegeschwindigkeit verlangsamt. Im zweiten Fall erklärten sie die Beschleunigung der Lesegeschwindigkeit mit der oberflächlicheren Verarbeitung des Satzes: Die Probanden stützen sich in der Verarbeitung auf die Semantik des Satzes, die syntaktische Struktur wird nur oberflächlich verarbeitet. Dies ist für die allgemein schnellere Lesegeschwindigkeit im Vergleich zur Behaltensaufgabe verantwortlich. Daneben wird bei der Verarbeitung zum Behalten in höherem Maße als bei der Verarbeitung zum Verständnis redundante Information berücksichtigt. Die Beschleunigung der Lesegeschwindigkeit erklären die Autoren mit der Begrenzung des semantischen Feldes: Das Ende des Satzes wird mehr und mehr vorhersehbar, je näher man ihm kommt. Damit beschleunigt sich die Verarbeitung.

Aaronson und Scarborroughs (1976) Erklärung für die Verlangsamung der Lesegeschwindigkeit ist im Einklang mit den oben vorgestellten Ideen von Just und Carpenter (1992) und Cowan (z.B. 1999). Auch innerhalb des Modells von Rummer und Schweppe (2009, s. auch Schweppe, 2006) lässt sich die Verlangsamung der Lesegeschwindigkeit mit einem Überschreiten der Kapazitäten der Aufmerksamkeit erklären. Die Aufmerksamkeit ist begrenzt und mit ihrer zunehmenden Belastung durch die Oberflächeninformation, die durch kontrollierte Prozesse aktiv gehalten werden muss, verlangsamt sich die Aufnahme und Aktivierung neu hinzu kommender Information. Beim Verstehen von Sätzen hingegen wird die Aktivierung ausschließlich zur Erhaltung und Integration des propositionalen Inhalts benötigt, wodurch die Aufmerksamkeit weniger belastet wird.

Da beim inhaltlichen Satzbehalten, ähnlich wie beim Verstehen, die Oberflächeninformation, also die genaue Wortwahl, syntaktische Struktur und sich daraus ergebende morphologische und phonologische Eigenschaften eines Satzes, nicht aktiv gehalten werden muss, wäre denkbar, dass diese Aufgabe, ähnlich wie das Verstehen von Sätzen, das Arbeitsgedächtnis in geringerem Maße belastet. Wäre dies der Fall, so könnte diese Modifikation der Aufgabe (inhaltliche Wiedergabe statt wörtliche) die Aufgabe auch

für Erhebungen in einer L2 und den Vergleich zwischen Mutter- und Zweitsprachlern geeignet machen. In Studie 4 soll anhand des Paradigmas von Aaronson und Scarborough (1976) getestet werden, ob die Belastung des Arbeitsgedächtnisses beim inhaltlichen Behalten von Sätzen geringer ausfällt als beim wörtlichen Behalten. Diese Belastung wird operationalisiert durch die Lesegeschwindigkeit (als mittlere Lesezeit pro Wort) und zunächst nur für die muttersprachliche Verarbeitung geprüft. Drei Bedingungen werden getestet: das wörtliche Behalten, das inhaltliche Behalten und das Verstehen. Es wird erwartet, dass sich die Lesegeschwindigkeiten in den drei Bedingungen unterscheiden. Da das wörtliche Behalten unter diesen drei Bedingungen die größte Belastung des Arbeitsgedächtnisses darstellt, wird für das wörtliche Behalten die niedrigste Lesegeschwindigkeit erwartet. Für das Lesen zum Verstehen wird, analog zu Aaronson und Scarborough (1976), eine hohe Lesegeschwindigkeit erwartet, da diese Bedingung nicht mit einer Behaltensleistung einhergeht. Das inhaltliche Behalten fordert zwar eine Behaltensleistung, jedoch ist diese auf den Inhalt beschränkt und sollte damit geringer ausfallen als beim wörtlichen Behalten. Daher wird für das inhaltliche Behalten eine mittlere Geschwindigkeit erwartet. Daneben wird, ebenfalls analog zu den Ergebnissen von Aaronson und Scarborough (1976), erwartet, dass sich die Veränderung der Lesegeschwindigkeit über den Satz hinweg in den Bedingungen unterscheidet. Dabei wird eine Verlangsamung der Lesegeschwindigkeit beim wörtlichen Behalten (höchste Belastung des Arbeitsgedächtnisses und damit auch über den Satz hinweg höchster Anstieg der Belastung), eine geringere Verlangsamung beim inhaltlichen Behalten (mittlere Belastung und damit mittlerer Anstieg der Belastung) und eine Beschleunigung der Lesegeschwindigkeit beim Verstehen erwartet (analog zu Aaronson und Scarborough, 1976). Dies soll sich in einer Interaktion der Lesegeschwindigkeit an zwei Positionen des Satzes (Satzanfang - Satzende) mit der Instruktion (wörtliches Behalten - inhaltliches Behalten - Verstehen) zeigen. Um sicherzugehen, dass sich die beiden Behaltensbedingungen voneinander (und nicht nur von der Verstehensbedingung) unterscheiden, sollen diese außerdem separat betrachtet werden. Die Interaktion zwischen den Faktoren soll auch bei dieser separaten Analyse der beiden Behaltensbedingungen zu finden sein.

9.1 Methode

Stichprobe

90 Studenten der Universität Erfurt im Alter von 19 bis 22 Jahren (73 weiblich) nahmen gegen Versuchspersonenstunden oder Bezahlung an der Studie teil. Die Muttersprache aller Teilnehmer ist Deutsch. Die Probanden gaben im Vorfeld schriftlich ihr Einverständnis zur Teilnahme.

Material

Für das Experiment wurden jeweils die schriftlichen Versionen der 40 adaptierten Sätze aus der deutschen Version des Lesespannentests verwendet, die bereits in Studie 2 eingesetzt wurden. Für die Bedingung "Verstehen" wurde zu jedem Satz eine Verständnisfrage formuliert, die mit "ja" oder "nein" zu beantworten war (s. Anhang A5).

Ablauf

Die Sätze wurden computergesteuert am Bildschirm präsentiert. Die Präsentation erfolgte Wort für Wort und wurde von den Teilnehmern selbst durch das Drücken der Leertaste gesteuert. Dabei wurden die Abstände zwischen dem Drücken der Taste registriert; diese werden im Folgenden als Lesezeit pro Wort interpretiert. Dieser Ablauf wurde aus dem Paradigma von Aaronson und Scarborough (1976) übernommen. Die Wörter wurden im Zentrum des Bildschirms in schwarzer Schrift auf weißem Bildschirm und in Arial, Größe 22 präsentiert. Zu Beginn jedes Satzes wurde ein Stern angezeigt, um die Aufmerksamkeit auf die Bildschirmmitte zu lenken. Die Teilnehmer wurden zufällig in drei gleich große Gruppen mit unterschiedlicher Instruktion aufgeteilt (s. Anhang C1). Gruppe 1 wurde angewiesen, die Sätze nach der Präsentation wörtlich wiederzugeben, Gruppe 2 sollte die Sätze inhaltlich wiedergeben. Die Wiedergabe erfolgte schriftlich auf vorbereiteten Antwortbögen. Nach dem letzten Wort eines Satzes wurde jeweils ein leerer weißer Bildschirm angezeigt und die Teilnehmer hatten Zeit, die Sätze niederzuschreiben. Durch Drücken der Leertaste konnten die Probanden den nächsten Durchgang starten. In Gruppe 3 wurden die Teilnehmer instruiert, die Sätze zu lesen und jeweils eine Verständnisfrage zu beantworten. Im Anschluss an jeden Satz wurde eine Frage im Zentrum des Bildschirms angezeigt, die die Teilnehmer durch Druck auf die rechte bzw.

linke Umschalttaste mit „Ja“ bzw. „Nein“ beantworten sollten; die Tasten waren entsprechend markiert worden. Die Beantwortung der Frage startete den nächsten Durchgang. Drei Übungsdurchgänge machten die Teilnehmer mit dem jeweiligen Ablauf vertraut. Die Sitzungen dauerten ca. 45 Minuten in den Gruppen 1 und 2 und ca. 20 Minuten in Gruppe 3.

Auswertung

Für jeden Satz wurde zunächst innerhalb der Bedingungen die mittlere Lesedauer pro Wort ermittelt, dieser Wert wird im Folgenden als Lesegeschwindigkeit bezeichnet. Um Verzerrungen der Daten zu vermeiden, wurden das jeweils erste und letzte Wort jedes Satzes nicht in die Berechnung mit einbezogen: Das letzte Wort eines Satzes bzw. das erste Wort des Folgesatzes waren teilweise auf dem Bildschirm noch angezeigt, während die Probanden mit der Wiedergabe des Satzes beschäftigt waren. Dies war eine Folge des schnellen und damit gelegentlich vorschnellen Tastendrucks. Da die Sätze zwischen 16 und 20 Wörtern lang sind, hat der Ausschluss des jeweils ersten und letzten Wort eines jeden Satzes keinen Einfluss auf die Hypothesenprüfung. Daneben wurde jeweils (ebenfalls innerhalb der Bedingungen) die mittlere Lesedauer der 4 ersten (Satzbeginn) und 4 letzten Wörter (Satzende) eines jeden Satzes gebildet, dabei wurden wiederum das erste und letzte Wort des Satzes nicht berücksichtigt. Es wurden also für den Satzanfang die Wörter an den Positionen 2, 3, 4 und 5 mit einbezogen, für das Satzende entsprechend das fünftletzte bis zweitletzte Wort. Der Abstand von Satzanfang und Satzende variierte je nach Satzlänge zwischen 6 und 10 Wörtern. Die Sätze wurden nur in Abhängigkeit ihrer Zugehörigkeit zu den Bedingungen, nicht in Abhängigkeit von den Probanden betrachtet. Es fließen also in jeder Bedingung die Werte aus 1200 Sätzen in die Auswertung ein.

9.2 Ergebnisse

Tabelle 8: Mittlere Lesezeiten pro Wort in den drei Gruppen zu Satzbeginn und Satzende.

Gruppe	Satzbeginn	Satzende
wörtl. Wiedergabe	618,27 ms (SD = 527,03)	770,71 ms (SD = 971,77)
inhaltl. Wiedergabe	544,69 ms (SD = 350,12)	610,25 ms (SD = 513,19)
keine Wiedergabe	463,57 ms (SD = 185,67)	446,28 ms (SD = 165,17)

Lesegeschwindigkeit pro Wort

Die mittleren Lesegeschwindigkeiten pro Wort unterscheiden sich signifikant zwischen den drei Gruppen, $F(1,3597) = 163,97$; $p < .001$. Es ergaben sich mittlere Lesezeiten von 785,05 ms/Wort (SD = 596,67; $n = 1200$) für das wörtliche Behalten, 640,71 ms/Wort (SD = 434,17; $n = 1200$) für das inhaltliche Behalten und 463,72 ms/Wort (SD = 155,41; $n = 1200$) für das Verstehen.

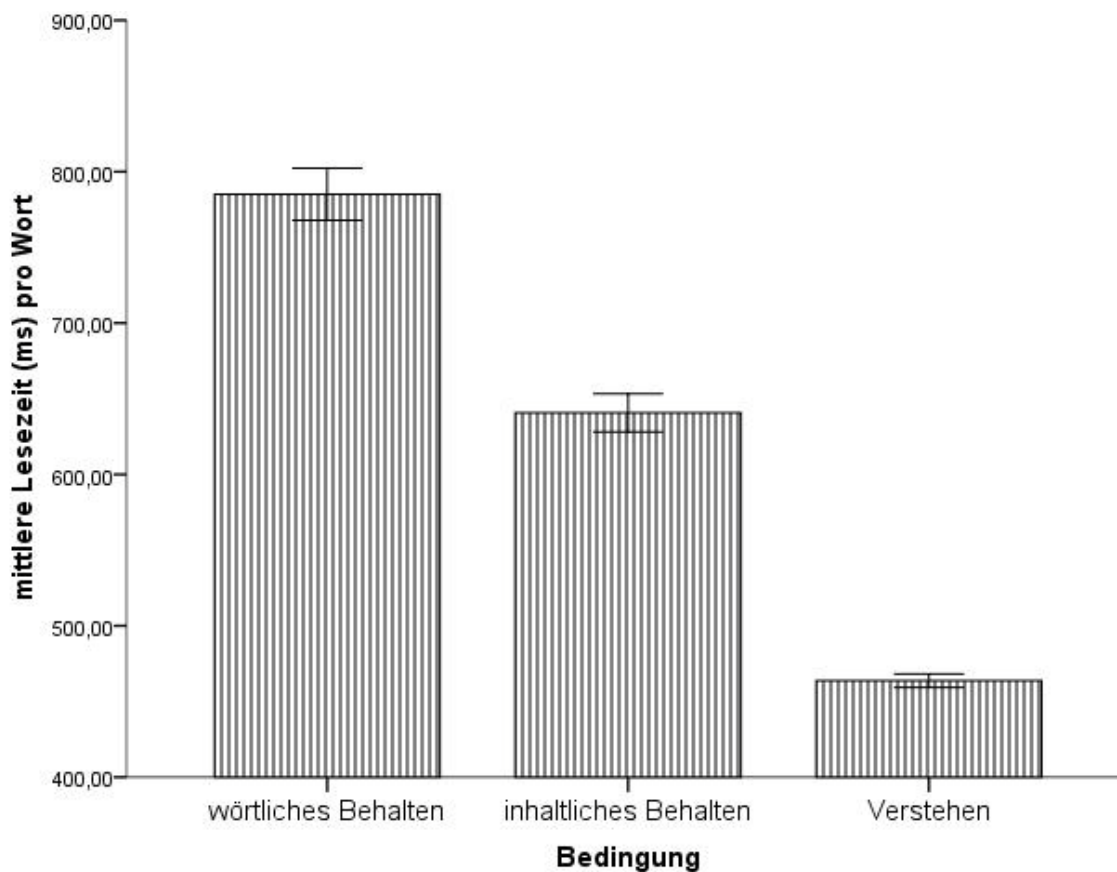


Abbildung 12: Vergleich der mittleren Lesezeit pro Wort (in ms) zwischen den drei Bedingungen wörtliches und inhaltliches Behalten und Verstehen von Sätzen.

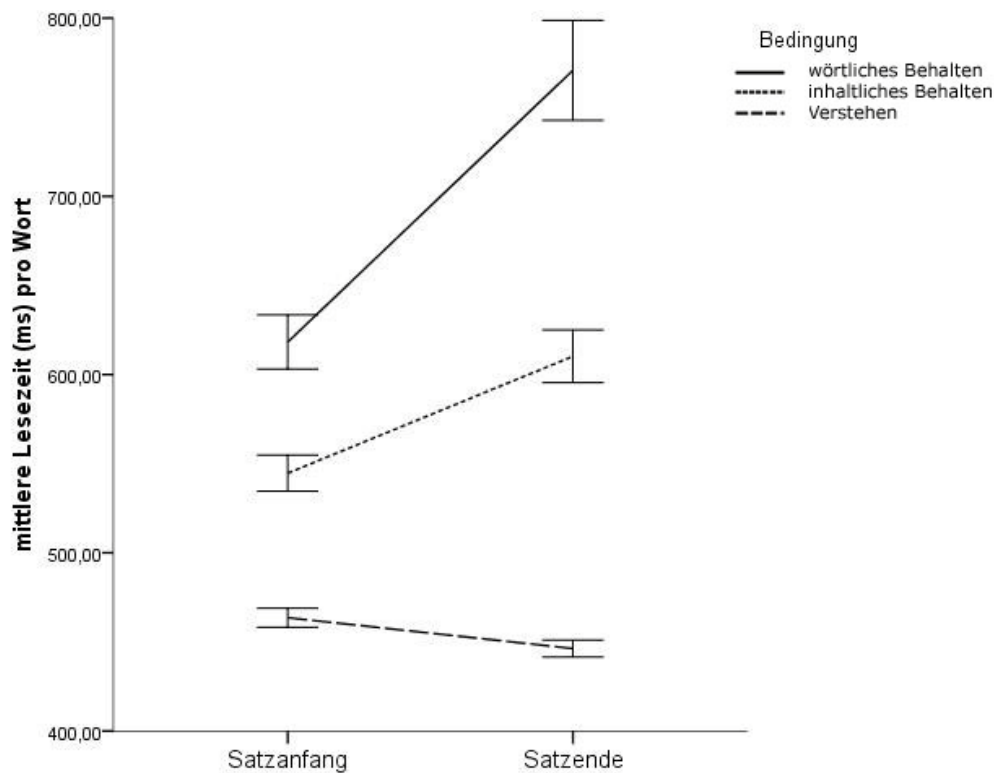


Abbildung 13: Vergleich der mittleren Lesezeit pro Wort (in ms) zwischen den drei Bedingungen wörtliches und inhaltliches Behalten und Verstehen von Sätzen zum Satzanfang und Satzende.

Geplante Vergleiche zeigen, dass alle Gruppen sich voneinander unterscheiden (Gruppen 1 und 2: $F(1,3597) = 65,94$; $p < .001$; Gruppen 2 und 3: $F(1,3597) = 99,16$; $p < .001$; Gruppen 1 und 3: $F(1,3597) = 326,82$; $p < .001$; vgl. Abbildung 12).

Lesegeschwindigkeit und Position im Satz

Tabelle 8 zeigt die mittleren Lesegeschwindigkeiten in den drei Gruppen zu Satzbeginn und Satzende. Eine ANOVA mit Messwiederholung und den Faktoren POSITION (Satzbeginn vs. Satzende) und INSTRUKTION (wörtl. Satzbehalten vs. inhaltl. Satzbehalten vs. Verstehen) ergaben sich signifikante Haupteffekte für beide Faktoren (POSITION: $F(1,3597) = 44,91$; $p < .001$; INSTRUKTION: $F(1,3597) = 91,31$; $p < .001$). Die 2-fach Interaktion erreicht ebenfalls die Signifikanzgrenze ($F(1,3597) = 24,09$; $p < .001$; vgl. Abbildung 13).

Geplante Einzelvergleiche ergaben einen Unterschied zwischen Satzbeginn und Satzende für die wörtliche Behaltensinstruktion, $F(1,3697) = 77,72$; $p < .001$, und die

inhaltliche Behaltensinstruktion, $F(1,3597) = 14.38$; $p < .001$, jedoch keinen Unterschied zwischen den beiden Positionen in der Verstehensbedingung, $F < 1$.

Als letzter Schritt soll aufgezeigt werden, dass sich die Lesegeschwindigkeit zwischen den beiden Behaltensaufgaben über den Satz hinweg unterschiedlich verhält. Es ist notwendig, dies in einer separaten Analyse zu zeigen, da die Interaktion in der ANOVA mit allen drei Bedingungen auf die Unterschiede zwischen der Verstehensbedingung und den beiden Behaltensbedingungen zurückzuführen sein könnte.

Wurden in die ANOVA mit Messwiederholung nur die beiden Behaltensbedingungen einbezogen, so ergaben sich ebenfalls Haupteffekte für POSITION ($F(1,2398) = 54,25$; $p < .001$) und INSTRUKTION ($F(1,2398) = 30,36$; $p < .001$) und eine Interaktion zwischen POSITION und GRUPPE ($F(1,2398) = 8,62$; $p < .005$).

9.3 Diskussion

In diesem Experiment sollte geprüft werden, ob das inhaltliche Behalten eines Satzes bei Muttersprachlern mit einer geringeren Belastung des Arbeitsgedächtnisses einhergeht als das wörtliche Behalten. Dazu wurde das Paradigma einer Lesestudie von Aaronson und Scarborough (1976) adaptiert und um die Bedingung "inhaltliches Behalten" erweitert. Die Belastung des Arbeitsgedächtnisses während des Lesens wurde durch die Lesegeschwindigkeit (über den Satz gemittelte Lesedauer pro Wort) operationalisiert. Für das inhaltliche Behalten eines Satzes wurde eine Lesegeschwindigkeit gemessen, die höher war als die Lesegeschwindigkeit für das wörtliche Behalten und geringer als die Lesegeschwindigkeit für das Verstehen eines Satzes ohne Behaltensaufgabe. Dies bestätigt die Annahme einer geringeren Belastung des Arbeitsgedächtnisses beim inhaltlichen Behalten im Vergleich zum wörtlichen Behalten, zumindest in einer L1. Soll statt Inhalt und Satzoberfläche nur der propositionale Gehalt eines Satzes behalten werden, so erhöht sich bei Muttersprachlern die mittlere Lesegeschwindigkeit pro Wort, was als niedrigere Belastung des Arbeitsgedächtnisses interpretiert werden kann. Da jedoch auch beim inhaltlichen Behalten Informationen scheinbar in höherem Maße aktiv gehalten werden als beim Verstehen, liegt die Lesedauer pro Wort weiterhin über der Lesedauer pro Wort beim Lesen ohne Behaltensaufgabe. Vergleicht man die Lesegeschwindigkeit zu Satzbeginn und -ende, so zeigt sich, dass sich die Lesegeschwindigkeit in beiden Behaltensaufgaben (wörtlich und inhaltlich) zum Satzende hin verlangsamt (in Übereinstimmung mit den

Ergebnissen von Aaronson und Scarborough, 1976), jedoch konnte die von Aaronson und Scarborough (1976) festgestellte Beschleunigung beim Lesen zum Verstehen nur numerisch, aber nicht statistisch repliziert werden. Wichtiger für die Fragestellung dieser Arbeit ist jedoch, dass die Lesegeschwindigkeit mit der Instruktion „wörtliches Behalten“ in stärkerem Maße abnimmt, als mit der Instruktion „inhaltliches Behalten“, was ebenso wie die allgemein niedrigere Lesedauer pro Wort auf eine geringere Belastung der Aufmerksamkeitsressource schließen lässt.

Die Ergebnisse aus Studie 4 deuten darauf hin, dass das Lesen zum inhaltlichen Behalten für Muttersprachler eine geringere Belastung des Arbeitsgedächtnisses mit sich bringt als das Lesen zum wörtlichen Behalten. Sollte dies auch für *Near-Natives* so sein, so sollte sich die Leistung von Muttersprachlern und *Near-Natives* in einer inhaltlichen Behaltensaufgabe annähern im Vergleich zur wörtlichen Behaltensaufgabe. Dies soll in Studie 5 getestet werden.

Konkret soll geprüft werden, ob die mehrmals aufgezeigte Interaktion von Aufgabentyp und Sprachhintergrund (L1 oder L2 Deutsch) weiter Bestand hat, wenn statt der wörtlichen Wiedergabe die Leistung in der inhaltlichen Satzwiedergabe in einer L2 betrachtet wird. Dazu wird die wörtliche und inhaltliche Wiedergabe von Sätzen bei Muttersprachlern und *Near-Natives* getestet und jeweils zu den Leistungen der Teilnehmer im C-Test in Bezug gesetzt. Es wird erwartet, dass in den zum wörtlichen Behalten instruierten Gruppen der in Studie 2 aufgezeigte Effekt (niedrigere Leistung in der Satzwiedergabe in einer Zweitsprache als auf Grundlage der allgemeinen Sprachkompetenz und der Leistung von Muttersprachlern zu erwarten ist) repliziert werden kann. Für die L1- und L2-Gruppen, die instruiert werden, die Sätze inhaltlich wiederzugeben, soll sich eine Annäherung der Behaltensleistung zeigen. Statistisch würde sich dies in einer Interaktion der Faktoren TEST (C-Test - Satzwiedergabe) und L1 (Deutsch L1 - Deutsch L2) im ersten Fall und der Abwesenheit einer Interaktion dieser Faktoren im zweiten Fall zeigen.

10 Studie 5 - Wörtliches und inhaltliches Satzbehalten in L1 / L2 Deutsch

In Studie 5 wird untersucht, ob die geringere Belastung durch die Einschränkung der Behaltensanforderung auf den Satzinhalt zu höherer Leistung in der inhaltlichen Satzwiedergabe führt. Insbesondere soll betrachtet werden, ob diese Modifikation der Aufgabenstellung dazu führt, dass sich die Leistung von Muttersprachlern und *Near-Natives* angleicht. Dies würde darauf hindeuten, dass tatsächlich die hohe Beanspruchung der Aufmerksamkeit, die mit der Aufrechterhaltung der Information zur Satzoberfläche einhergeht für das schlechtere Abschneiden der *Near-Natives* verantwortlich ist. Dazu wird die Leistung von Muttersprachlern und *Near-Natives* bei der inhaltlichen und wörtlichen Wiedergabe von Sätzen erhoben und mit ihrer im C-Test erfassten allgemeinen Sprachkompetenz in Bezug gesetzt. In zwei 2x2 Designs mit den Faktoren TEST (C-Test vs. Satzwiedergabe) und L1 (L1 Deutsch vs. L2 Deutsch) werden die Gruppen getrennt nach Instruktionsbedingung (wörtliche Satzwiedergabe und inhaltliche Satzwiedergabe) analysiert. Es wird erwartet, dass die in den vorangegangenen Studien gefundene Interaktion zwischen TEST und L1 für die beiden Gruppen gefunden wird, die instruiert werden, die Sätze wörtlich wiederzugeben, jedoch nicht für die Gruppen, die instruiert werden, die Sätze inhaltlich wiederzugeben.

10.1 Methode

Stichprobe

44 Studenten an verschiedenen Universitäten in Berlin und der Universität Erfurt wurden für die Studie rekrutiert. Davon sprechen 22 (16 weiblich) Deutsch als Muttersprache, 22 (12 weiblich) sind *Near-Natives*, also hochkompetente Sprecher mit sehr frühem Spracherwerb ihrer L2 Deutsch. Die Muttersprachler waren zum Erhebungszeitpunkt durchschnittlich 22,35 Jahre alt ($SD = 2,63$, Range: 18 - 27 Jahre), die *Near-Natives* 22,76 Jahre ($SD = 2,08$, Range: 19 - 27 Jahre). Alle *Near-Natives* sprechen Deutsch ohne Akzent und leben seit durchschnittlich 16,57 Jahren ($SD = 5,18$; Range: 5 - 24 Jahre) in Deutschland, sie lernten Deutsch durchschnittlich ab dem Alter von 6,14 ($SD = 3,17$, Range: 3 - 13 Jahre). Sie haben in Deutschland das Abitur gemacht und besuchten zum Zeitpunkt der Erhebung eine deutsche Universität. Die Erstsprachen der Teilnehmer sind Russisch (10 TN), Arabisch (2 TN), Italienisch (2 TN), Bulgarisch (2 TN), Türkisch

(2 TN), Litauisch (1 TN), Französisch (1 TN), Griechisch (1 TN) und Niederländisch (1 TN).

Material

20 Sätze aus den Studien 2, 3 und 4 wurden für Studie 5 verwendet (s. Anhang A6). Daneben wurde der deutsche C-Test aus den Studien 2 und 3 eingesetzt.

Ablauf

Die 20 Sätze wurden in randomisierter Reihenfolge auditiv über Kopfhörer präsentiert. Die Teilnehmer wurden zufällig den beiden Bedingungen zugeteilt und in einem ruhigen Raum am Computer getestet. In Erfurt fanden die Versuche im psychologischen Labor statt, in Berlin wurde tageweise ein Büroraum angemietet. Jeweils die Hälfte der Teilnehmer in jeder Sprachgruppe (L1 Deutsch bzw. L2 Deutsch) wurde angewiesen, die Sätze wörtlich wiederzugeben (Instruktion "wörtlich", vgl. Anhang C), die andere Hälfte sollte bei der Wiedergabe nur den propositionalen Gehalt berücksichtigen, Wortwahl und Syntax waren in dieser Bedingung nicht relevant (Instruktion "inhaltlich", vgl. Anhang C). Die Teilnehmer wurden schriftlich am Computer instruiert und angewiesen, der Testleiterin die Instruktion nach dem Lesen in eigenen Worten wiederzugeben. Die Wiedergabe erfolgte schriftlich auf vorbereiteten Antwortbögen, vorab absolvierten die Teilnehmer zwei Übungsdurchgänge, um sich mit dem Ablauf vertraut zu machen. Im Anschluss bearbeiteten alle Teilnehmer den C-Test und füllten einen Fragebogen zu ihrer Spracherwerbsgeschichte aus.

Auswertung

Aus den zwischen den Teilnehmern gekreuzten Variablen L1 und INSTRUKTION ergeben sich vier Gruppen: L1 Deutsch und wörtliche Instruktion (Gruppe L1-W), L1 Deutsch und inhaltliche Instruktion (Gruppe L1-I), L2 Deutsch und wörtliche Instruktion (L2-W) und L2 Deutsch und inhaltliche Instruktion (Gruppe L2-I). Ein Vergleich der beiden Gruppen mit L2 Deutsch zeigte keine Unterschiede in ihrem durchschnittlichen L2-AoA (Gruppe L2-W: $M = 6,20$ Jahre, $SD = 3,20$, Range: 3 - 12 Jahre; Gruppe L2-I: $M = 6,3$ Jahre, $SD = 3,27$, Range: 3 - 13 Jahre) und nur eine geringe Abweichung in ihrer Aufenthaltsdauer in Deutschland von durchschnittlich 17,7 Jahren ($SD = 4,75$, Range: 8 -

24 Jahre) in Gruppe L2-W und 15,2 Jahren (SD = 5,73, Range: 5 - 24 Jahre) in Gruppe L2-I.

Die Daten aller Teilnehmer wurden ungeachtet der Instruktion auf zwei Arten ausgewertet, nämlich gemäß Kriterien der wörtlichen und der inhaltlichen Wiedergabe. Die erste Auswertung der Daten gemäß den Kriterien der wörtlichen Wiedergabe erfolgte analog zur Auswertung in den Studien 2 und 3: Es wurde die Anzahl der korrekt wiedergegebenen Wörter pro Satz gezählt. Dieser Wert wurde dann in Bezug zur Wortanzahl des Satzes gesetzt und in Prozentwerte überführt (Prozent der Wörter des Satzes, die korrekt wiedergegeben wurden). Die Auswertung gemäß des Satzinhaltes erfolgte gemäß des von Brewer, Sampaio und Barlow (2005) entwickelten Schemas zur Beurteilung der Wiedergabe des propositionalen Gehalts eines Satzes, das im Folgenden beschrieben wird. Die Wiedergabe eines jeden Satzes wurde einer von fünf Kategorien zugeordnet: korrekt, Paraphrase, teilweise fehlerhaft, fehlerhaft, nicht vorhanden. In die Kategorie "korrekt" fielen alle Sätze, die genau dem vorgegebenen Satz entsprachen, mit Ausnahme von Fehlern in der Rechtschreibung oder gängigen Abkürzungen. Als "Paraphrase" wurden Sätze gewertet, die synonyme Wortersetzungen enthielten oder alternative Formulierungen, die als Paraphrase betrachtet werden können und solche, in denen redundante Information weggelassen oder implizit vorhandene Information hinzugefügt wurde. "Teilweise fehlerhaft" waren solche Reproduktionen, die einige Informationen des Satzes enthielten, aber nicht genug, um als Paraphrase zu gelten. Dazu gehörten Fehler wie das nicht-synonyme Ersetzen von Wörtern, das Weglassen nicht-redundanter Information und das Hinzufügen nicht implizierter Information. "Fehlerhaft" waren Reproduktionen, die keine inhaltliche Information des Originalsatzes enthielten und als "nicht vorhanden" wurden Sätze gewertet, die gar nicht wiedergegeben worden waren. Für die Beurteilung der inhaltlichen Wiedergabeleistung wurden zwei unabhängige Bewerter herangezogen, deren Einordnung der Sätze in die Kategorien eine sehr hohe Übereinstimmung aufwies (Cronbachs $\alpha = 0,937$). Die beiden Bewerter waren 30 Jahre alt und weiblich, bzw. 32 Jahre alt und männlich. Sie sind Muttersprachler des Deutschen und in Deutschland aufgewachsen. Für jedes Rating wurde zunächst für jeden Probanden berechnet, wie viel Prozent der Sätze in die jeweiligen Kategorien fiel (analog zum Vorgehen von Brewer et al., 2005). Dabei wurde aus der Summe der Sätze, die den Kategorien "korrekt" oder "Paraphrase" zugeordnet waren, ein Maß für die inhaltliche Wiedergabeleistung gebildet (vgl. Brewer et al., 2005). Im letzten Schritt wurden diese Prozentwerte über die beiden Ratings gemittelt, so dass beide Bewertungen in die Analyse

der Daten einfließen. Das Maß der inhaltlichen Wiedergabe ist also die Summe der als "korrekt" oder "Paraphrase" gewerteten Antworten (wiedergegebenen Sätze) als Prozent der Gesamtzahl der präsentierten Sätze, über das Urteil der beiden Bewerter gemittelt. Dieser Wert liegt den folgenden Berechnungen zu Grunde.

10.2 Ergebnisse

Tabelle 9: Mittlere Wiedergabeleistung in Prozent (SD in Klammern) und Leistung im C-Test für die Bedingungen und Sprachgruppen²¹.

	C-Test	wörtliche Wiedergabe	inhaltliche Wiedergabe
L1 - wörtlich	95 (2,57)	89,48 (3,74)	92,27 (5,96)
L1 - inhaltlich	95,1 (2,55)	86,57 (6,18)	85,23 (11,91)
L2 - wörtlich	96,45 (2,07) [96,3 (2,11)]	86,4 (6,63) [85,29 (5,81)]	82,73 (10,87)
L2 - inhaltlich	94,91 (2,63)	73,87 (15,02)	66,59 (11,79)

²¹ Für die Gruppe L2 wörtlich sind zusätzlich in eckigen Klammern die Daten unter Ausschluss einer VP aufgeführt. Diese VP zeigt eine außergewöhnlich hohe Leistung in der Satz wiedergabe, es scheint daher interessant, die Daten zusätzlich unter Ausschluss dieser VP zu berichten.

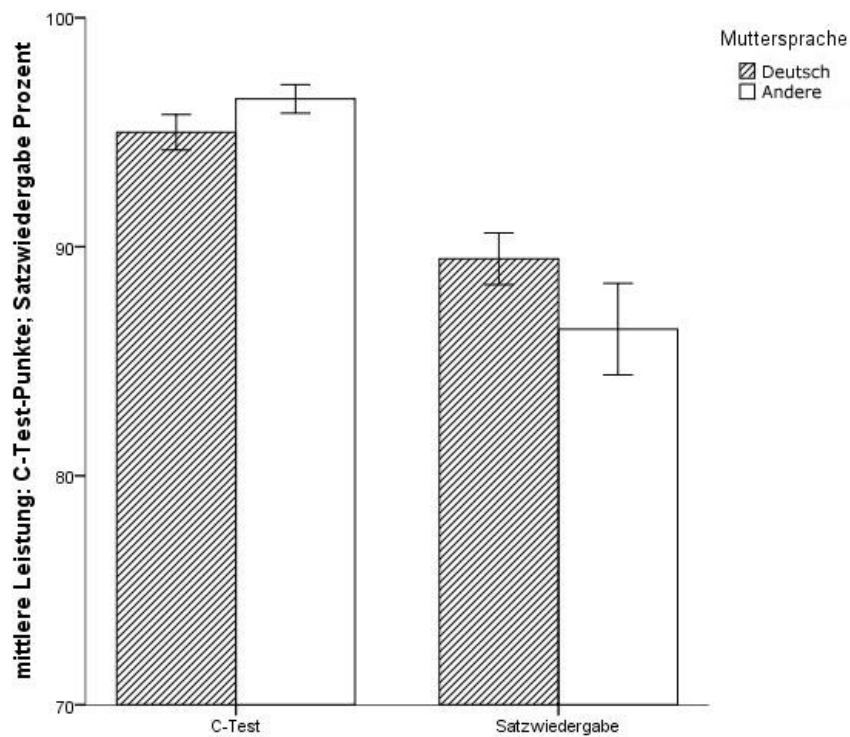


Abbildung 14: Vergleich der erreichten Punktzahl im C-Test und der wörtlichen Satzwiedergabe für die Gruppen mit der Instruktion zur wörtlichen Satzwiedergabe.

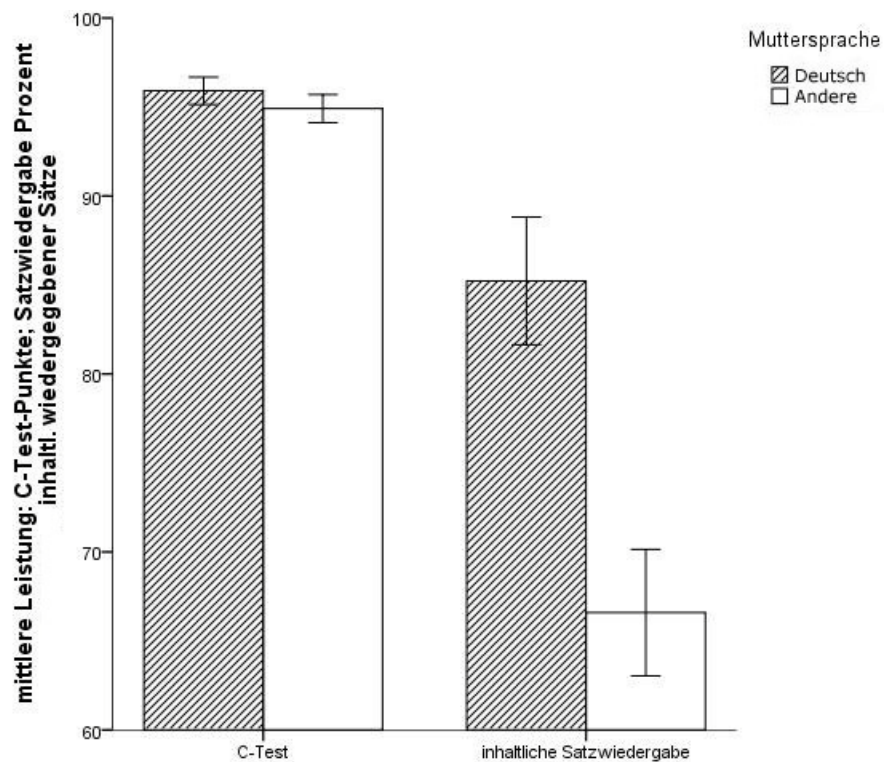


Abbildung 15: Vergleich der erreichten Punktzahl im C-Test und der inhaltlichen Satzwiedergabe (Kategorien "korrekt" und "Paraphrase" summiert) für die Gruppen mit der Instruktion zur inhaltlichen Satzwiedergabe.

Die vier Gruppen unterscheiden sich in ihrer Leistung im C-Test nicht signifikant voneinander, weichen jedoch numerisch in ihrer mittleren Leistung ab. Gruppe L2-W (*Near-Natives*, wörtliche Wiedergabe) zeigte die beste durchschnittliche Leistung mit $M = 96,5$ Punkten ($SD = 2,07$) und die Gruppe L2-I (*Near-Natives*, inhaltliche Wiedergabe) die schlechteste Leistung mit $94,6$ Punkten ($SD = 2,63$). Die Gruppe L1-W (Muttersprachler, wörtliche Wiedergabe) lag mit $94,8$ Punkten ($SD = 2,57$) nur wenig über dieser Leistung, die Gruppe L1-I (Muttersprachler, inhaltliche Wiedergabe) schnitt mit $95,6$ Punkten ($SD = 2,55$) besser ab. Aufgrund der unterschiedlichen Auswertungsschemata in den beiden Bedingungen wörtliche und inhaltliche Wiedergabe werden die Daten im Folgenden für beide Bedingungen getrennt analysiert.

Ergebnisse wörtliche Wiedergabe

Die Muttersprachler zeigten eine numerisch bessere Wiedergabeleistung als die *Near-Natives*. (vgl. Tabelle 9) In einer ANOVA mit Messwiederholung und den Faktoren TEST (C-Test vs. wörtliche Wiedergabe) und L1 (L1 Deutsch vs. L2 Deutsch) ergab sich ein Haupteffekt für TEST, $F(1,20) = 66,69$; $p < .001$, aber kein Haupteffekt für L1, $F < 1$. Die Interaktion der beiden Faktoren erreichte die Signifikanzgrenze, $F(1,20) = 5,64$; $p < .05$. In geplanten Einzelvergleiche unterschieden sich die Gruppen in beiden Tests nicht signifikant (C-Test: $F(1,20) = 2,14$; Satzwiedergabe $F(1,20) = 1,79$).

In der Gruppe der *Near-Natives* mit der Instruktion zur wörtlichen Wiedergabe sind die Daten einer Versuchsperson auffällig hoch. Diese Person erreicht eine durchschnittliche Wiedergabeleistung von $97,51\%$, ein Wert, den kein Muttersprachler in dieser Stichprobe erreicht ($\max = 93,86\%$). Ihr C-Testwert liegt mit 98 Punkten ebenfalls sehr hoch. Dennoch sind die außergewöhnlich hohen Werte aus den Daten zur Spracherwerbsgeschichte ($AOA = 3$) nicht zu erklären und eine nachträgliche Befragung aufgrund der Anonymisierung der Daten ebenfalls nicht möglich. Dennoch sollen die Ergebnisse für die wörtliche Wiedergabe im Folgenden noch einmal unter Ausschluss der Daten dieser VP berichtet werden.

Ergebnisse wörtliche Wiedergabe - Ausschluss einer VP

Die Daten für Gruppe L2-W unter Ausschluss der VP sind in Tabelle 9 in eckigen Klammern aufgeführt. Eine ANOVA mit Messwiederholung ergibt einen Haupteffekt für

TEST, $F(1,19) = 90,82$; $p < .001$, aber keinen Haupteffekt für L1 ($F < 1$). Die Interaktion zwischen den Faktoren wird signifikant, $F(1,19) = 9,99$; $p = .005$. Geplante Einzelvergleiche zeigen keinen Unterschied zwischen den Gruppen in ihrer C-Testleistung ($F < 2$), aber einen marginal signifikanten Unterschied zwischen den Gruppen in der Satzwiedergabe, $F(1,19) = 3,93$; $p = .062$.

Ergebnisse inhaltliche Wiedergabe

In einer 2x2-ANOVA mit Messwiederholung und den Faktoren TEST (C-Test vs. inhaltliche Satzwiedergabe) und L1 (L1 Deutsch vs. L2 Deutsch) ergab sich ein Haupteffekt für TEST, $F(1,20) = 59,56$; $p < .001$, und ein Haupteffekt für L1, $F(1,20) = 13,79$; $p < .001$. Auch die Interaktion zwischen den Faktoren wurde signifikant, $F(1,20) = 12,18$; $p < .01$. Geplante Einzelvergleiche ergaben außerdem einen Unterschied in der Wiedergabeleistung, $F(1,20) = 13,60$; $p = .001$, aber keinen Unterschied in der C-Testleistung, $F < 1$.

Da jedoch das Auswertungsschema für die inhaltliche Wiedergabe an Aussagen über Häufigkeiten gebunden ist (wie viel Prozent der wiedergegebenen Sätze entsprachen den Kategorien "korrekt" und "Paraphrase"), statt wie die beiden Verfahren zur wörtlichen Wiedergabe und zur C-Testbewertung jeweils Punkte pro korrekt wiedergegebenen / ergänztem Wort zu vergeben²², ist nicht sicher, ob die Analyse der inhaltlichen Wiedergabe (auch nach der im Auswertungsschema vorgesehenen Überführung der Häufigkeiten in Prozentwerte) sinnvoll durchgeführt werden kann. Zusätzlich soll daher noch eine Analyse der z-transformierten Daten berichtet werden. Dazu wurde für jede VP die Anzahl der wiedergegebenen Sätze in z-Werte übertragen, die entweder der Kategorie "korrekt" oder "Paraphrase" zugeordnet wurde²³. Daneben wurden die C-Testergebnisse für die beiden Gruppen für diese Analyse ebenfalls z-transformiert.

Ergebnisse inhaltliche Analyse - z-transformiert

Eine ANOVA mit Messwiederholung mit z-transformierten Werten und den Faktoren TEST (C-Test vs. inhaltliche Satzwiedergabe) und L1 (L1 Deutsch vs. L2

²² Bei der Bewertung der wörtlichen Satzwiedergabe wird die erreichte Wortanzahl in einen Prozentwert überführt, beim C-Test ist der erreichte Punktzahl näherungsweise am Prozentwert, ist sie doch eine Aussage der erreichten Punkte von 99. Jedoch beruhen beide Verfahren auf dem Zählen von einzelnen Wörtern und sind sich untereinander damit ähnlicher als dem kategorialen Auswertungsschema der inhaltlichen Auswertung.

²³ Dieser Wert wurde zuvor über beide Ratings gemittelt.

Deutsch) ergab einen Haupteffekte für L1, $F(1,20) = 7,84$; $p < .05$ und TEST ($F < 1$). Die Interaktion zwischen den beiden Faktoren wurde nicht signifikant, $F(1,20) = 2,49$; $p = .13$. Die Einzelvergleiche ergaben wie in der vorhergehenden Analyse einen Unterschied in der Wiedergabeleistung, $F(1,20) = 13,60$; $p = .001$, aber keinen Unterschied in der C-Testleistung, $F < 1$.

Vergleich der Instruktionsarten innerhalb der Sprachgruppen

In einem weiteren Schritt soll die Wirkung der Instruktion auf die wörtliche Wiedergabe innerhalb der Sprachgruppen betrachtet werden. Dazu werden auch die Daten der Gruppen mit der Instruktion "inhaltliches Behalten" gemäß der Methode der wörtlichen Auswertung (Wörter zählen, Prozentwert aus korrekt wiedergegebenen Wörtern pro Satz) bewertet. Es sollte also betrachtet werden, welcher Prozentsatz der Wörter unter beiden Instruktionsarten wörtlich wiedergegeben wurde und ob sich dieser Wert zwischen den Instruktionsbedingungen unterscheidet. Dies wurde für jede Sprachgruppe einzeln in einem t-Test mit unabhängigen Stichproben bestimmt. Dabei zeigte sich, dass die Gruppe der

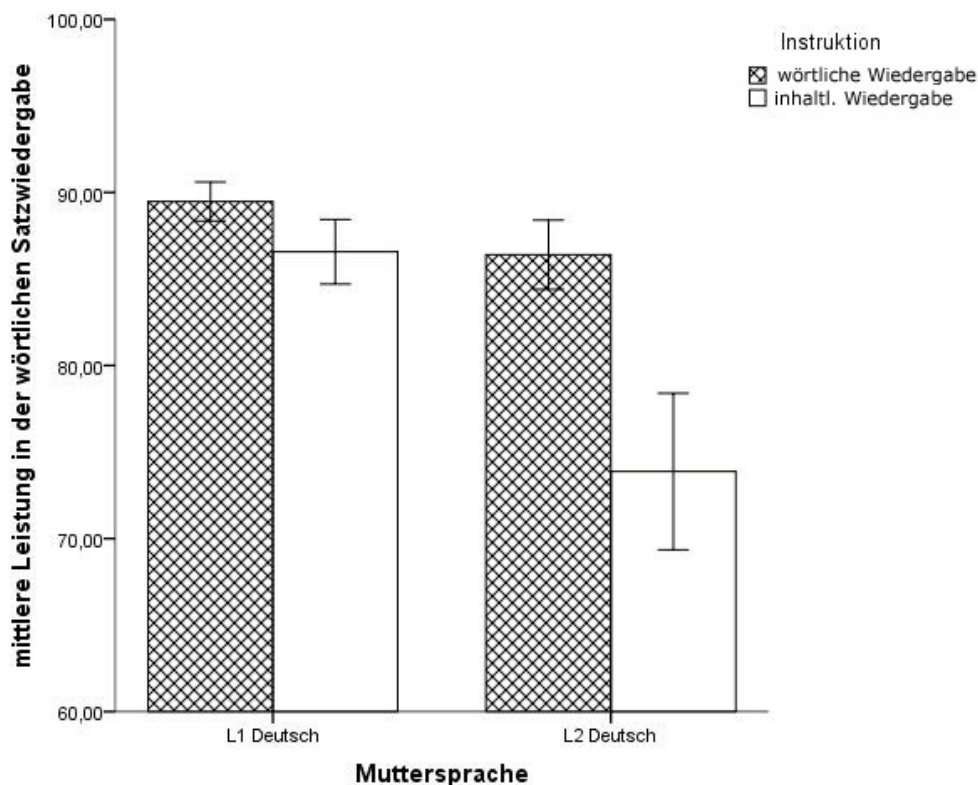


Abbildung 16: Vergleich der Leistung in der wörtlichen Satzwiedergabe zwischen den Sprachgruppen (L1 Deutsch vs. L2 Deutsch) und Instruktionsbedingungen (wörtliche vs. inhaltliche Satzwiedergabe).

Near-Natives, die zum inhaltlichen Behalten instruiert worden war, einen geringeren Prozentsatz der Wörter wörtlich wiederholte, als die Gruppe der *Near-Natives*, die zur wörtlichen Wiedergabe instruiert worden war, $t(df=20) = 2,53; p < .05$. Die beiden muttersprachlichen Gruppen (Instruktion zur wörtlichen Wiedergabe vs. Instruktion zur inhaltlichen Wiedergabe) jedoch unterschieden sich hinsichtlich der Wörter, die wörtlich wiedergegeben wurden, statistisch nicht voneinander, $t < 1,5$.

10.3 Diskussion

Das Ziel dieser Studie war es zu prüfen, ob die Beschränkung der Behaltensaufgabe auf den Inhalt des Satzes (statt Inhalt und Oberfläche) dazu führt, dass die Wiedergabeleistung der *Near-Natives* nicht mehr hinter der Leistung zurück bleibt, die aufgrund der C-Testleistung der *Near-Natives* und der Vergleichswerte der Muttersprachler zu erwarten wäre. Diese Frage kann durch die Ergebnisse von Studie 5 nicht abschließend beantwortet werden, da sich die Daten zur Leistung der *Near-Natives* und Muttersprachler mit der Instruktion zur inhaltlichen Satzwiedergabe widersprechen, je nachdem, ob man sie mit z-Transformation oder übertragen auf eine Prozentskala betrachtet. Während bei einer Umrechnung der Häufigkeitswerte in prozentuale Häufigkeiten eine Interaktion zwischen den Faktoren beobachtet werden kann, wird diese Interaktion nicht signifikant, wenn statt dessen die in z-Werte umgerechneten Daten zugrunde gelegt wird. Dennoch spricht auch der Trend, der sich in den Einzelvergleichen in den Analysen abzeichnet (sowohl bei Betrachtung der prozentualen Häufigkeiten, als auch bei Betrachtung der z-transformierten Werte), dagegen, dass die Leistung in der inhaltlichen Satzwiedergabe die Sprachkompetenz von *Near-Natives* besser widerspiegelt als die wörtliche Satzwiedergabe. Obwohl weitere Untersuchungen notwendig sind, um die Frage abschließend zu klären, scheint nach dieser ersten Betrachtung auch die inhaltliche Satzwiedergabe nicht geeignet, die Leistung von *Near-Natives* valide abzubilden.

In den beiden Gruppen, die instruiert wurden, die Sätze wörtlich wiederzugeben, konnte der bisher gefundene Effekt bestätigt werden. Es zeigte sich wieder eine Interaktion zwischen den Faktoren TEST und L1, welche auf das (statistisch nicht signifikante) schlechtere Abschneiden der *Near-Natives* in der wörtlichen Satzwiedergabe und ihre (ebenfalls nicht signifikante) bessere Leistung im C-Test zurückzuführen ist im Vergleich zur muttersprachlichen Vergleichsgruppe. Eine Versuchsperson der L2-Gruppe zeigte eine im Vergleich zu den bisherigen Ergebnissen ungewöhnlich hohe Leistung in der

Satzwiedergabe (97,51 %). Schließt man diese Versuchsperson aus der Analyse aus, so zeigen die übrigen Teilnehmer noch deutlicher den auch in den vorherigen Studien gefundenen Effekt: In den Einzelvergleichen zeigt sich nun, dass die Interaktion vermutlich auf den marginal signifikanten Unterschied in der Wiedergabeleistung zurückzuführen ist. Die Gruppen unterscheiden sich jedoch nicht im C-Test. Doch auch unter Einbezug der Daten dieser Versuchsperson bleibt die Interaktion der Faktoren stabil. Damit konnte der schon mehrmals aufgezeigte Effekt noch einmal repliziert werden.

Daneben zeigte sich in einer explorativen Analyse der wörtlichen Wiedergabe in beiden Instruktionsarten ein Zusammenhang zwischen der L1 und der Tendenz zur wörtlichen Wiedergabe der Sätze, auch bei der Möglichkeit zur inhaltlichen Wiedergabe. Teilnehmer mit L1 Deutsch gaben, auch wenn sie die Möglichkeit zur paraphrasierenden Wiedergabe hatten, die Sätze häufiger wörtlich wieder als Teilnehmer mit L2 Deutsch. Dies kann als Indiz dafür gedeutet werden, dass die wörtliche Satz-wiedergabe deutlich weniger aufwändig für die Muttersprachler ist: Sie wählen diese Wiedergabe auch wenn sie die Möglichkeit zur inhaltlichen Wiedergabe haben - womöglich ist es für sie einfacher, die Sätze wörtlich wiederzugeben. Dieser Befund ist im Einklang mit den Theorien von Potter und Lombardi (1990), die die Wiedergabe eines Satzes in seinem ursprünglichen Wortlaut als die präferierte Wiedergabe beschreiben. Die *Near-Natives*, die instruiert wurden, die Sätze inhaltlich wiederzugeben, geben die Sätze jedoch häufiger als die Muttersprachler in eigenen Worten wieder. Dies könnte darauf hindeuten, dass bei einer L2, anders als bei einer L1, der ursprüngliche Wortlaut bei der Wiedergabe nicht präferiert wird. Andererseits könnte dieser Unterschied aber auch auf eine unterschiedliche Interpretation der Instruktion zurückzuführen sein. Die *Near-Natives* sind sich durch Versuchsaufbau und Rekrutierung ihrer Eigenschaft als Nicht-Muttersprachler während des Experiments stark bewusst. Sie könnten daher besonderen Wert darauf legen, die Sätze in alternativen Formulierungen wiederzugeben, um damit ihre Sprachkompetenz zu demonstrieren, auch wenn es vielleicht für sie ebenfalls einfacher wäre, die Sätze wörtlich statt umformuliert wiederzugeben. Die Instruktion löst also speziell bei den *Near-Natives* Metakognitionen aus, die nicht nur Kapazitäten beanspruchen, sondern auch Restriktionen aufstellen (z.B. den selbst auferlegten "Zwang" zur Paraphrase), die die Leistung in der Aufgabe mindern. Diese Überlegungen sind spekulativ, machen jedoch die Schwierigkeiten, die durch die Instruktion zur inhaltlichen Wiedergabe hervorgerufen werden, deutlich.

Insgesamt zeigte sich in Studie 5, dass eine Modifikation der Aufgabe "Sätze wiederholen" ausgehend von der wörtlichen hin zur inhaltlichen Wiedergabe nicht zu einer besseren Vergleichbarkeit der Leistung von Muttersprachlern und *Near-Natives* führt. Dies ist ein weiterer Hinweis auf Unterschiede zwischen einer L2 und L1 in den Prozessen, die für die Leistung im Satzbehalten maßgeblich sind.

10.4 Gemeinsame Diskussion der Studien 4 und 5

In den Studien 4 und 5 wurde der Versuch unternommen, eine Alternative zur wörtlichen Satz wiedergabe zu identifizieren, die sich mit wenigen Änderungen der bestehenden Testverfahren umsetzen ließe. Außerdem sollte mit den Studien untersucht werden, ob sich mit der auf das inhaltliche beschränkten Satz wiedergabe, die eine geringere Belastung des Arbeitsgedächtnisses mit sich bringt, die Sprachkompetenz der *Near-Natives* mit größerer Validität abbilden lässt.

In Studie 4 konnte ich anhand der Lesegeschwindigkeit von Muttersprachlern demonstrieren, dass bei der Instruktion zum inhaltlichen Behalten von Sätzen das Lesen der Sätze mit einer geringeren Belastung der Aufmerksamkeit einhergeht als bei der Instruktion zum wörtlichen Behalten. Hierauf konnte aus gruppenspezifischen Unterschieden in der Lesegeschwindigkeit geschlossen werden. Diese sind ein Indiz dafür, dass das inhaltliche Behalten (zumindest bei Muttersprachlern) tatsächlich mit einer geringeren Belastung der Aufmerksamkeit einhergeht als das Behalten zur wörtlichen Wiedergabe.

Auf diesem Experiment aufbauend sollte in Studie 5 gezeigt werden, dass das aufgabenspezifisch schlechte Abschneiden der L2-Gruppe (im Vergleich zu C-Testleistung und der Leistung von Muttersprachlern) nur bei der wörtlichen Wiedergabe gefunden werden kann, aber nicht bei der inhaltlichen Wiedergabe. Während in der zur wörtlichen Satz wiedergabe instruierten Gruppe die Befunde aus dem ersten Teil dieser Arbeit noch einmal repliziert werden konnten, konnte diese zweite These nicht bestätigt werden. Trotz teilweise widersprüchlicher Daten in den beiden durchgeführten Analysen (Prozentskala und z-Transformation), scheint der aufgabenspezifische Nachteil auch bei der inhaltlichen Satz wiedergabe weiter zu bestehen.

Weiterhin erwies sich die Bedingung "inhaltliche Wiedergabe" als problematisch in Instruktion und Auswertung. Für die Auswertung der Ergebnisse sind mindestens zwei unabhängige Bewertungen notwendig, was den Einsatz der Aufgabe in der Praxis erschwert. Ein größeres Problem stellt jedoch die Instruktion dar: Auch wenn betont wird, dass beide Antwortmöglichkeiten, also die Wiedergabe des Satzes in den gleichen oder eigenen Worten, gleichberechtigt sind, werden Versuchspersonen womöglich einen Verdacht über die "erwünschte" Wiedergabe (wörtlich oder paraphrasiert) haben und versuchen, sich nach diesem zu richten. Insbesondere *Near-Natives*, die sich durch den Aufbau des Experiments und ihrer Rekrutierung als Zweitsprachler dieses Status' sehr bewusst sind, könnten sich herausgefordert fühlen, ihre Sprachkompetenz durch alternative Formulierungen zu demonstrieren. Dies hätte zur Folge, dass durch die Instruktion besonders bei Zweitsprachlern Metakognitionen ausgelöst werden, die nicht nur selbst einen Teil der kontrollierten Prozesse in Anspruch nehmen, sondern auch die Bearbeitung der Aufgabe durch die Restriktion zur Wiedergabe des Inhalts in anderen Worten erschweren. Diese selbst auferlegte Restriktion zur Wiedergabe in alternativer Wortwahl könnte für die schlechte Leistung der *Near-Natives* bei der inhaltlichen Satzwiedergabe verantwortlich sein. Der anzunehmende Vorteil durch die geringere Belastung des Arbeitsgedächtnisses könnte also durch die höhere Belastung durch Metakognitionen und/oder Restriktionen in der Wiedergabe moderiert werden. Die Leistung der *Near-Natives* in der inhaltlichen Satzwiedergabe ist also nur schwer zu interpretieren, weist aber andererseits auf praktische Probleme mit der Aufgabenstellung hin, die sie für den Einsatz in Sprachtests ungeeignet macht.

Die muttersprachlichen Teilnehmer in Studie 5 unterscheiden sich nur numerisch, aber nicht inferenzstatistisch in ihrer wörtlichen Wiedergabeleistung. Die Tendenz zur wörtlichen Wiedergabe der Sätze scheint für die Muttersprachler nicht abhängig von der Instruktion zu sein. Andererseits zeigt sich in Studie 4 zwischen den Gruppen mit den Instruktionen zum inhaltlichen und zum wörtlichen Behalten ein Unterschied in der Lesegeschwindigkeit, der als niedrigere Belastung des Arbeitsgedächtnisses während des Lesens zum inhaltlichen Behalten im Vergleich zum wörtlichen Behalten interpretiert wird. Die besondere Aufmerksamkeit, die beim Lesen zur wörtlichen Wiedergabe auf die Satzoberfläche gerichtet ist, scheint sich also nicht in besserer Behaltensleistung niederschlagen. Dieses Ergebnis ist überraschend, muss aber vielleicht ebenfalls mit Einschränkung interpretiert werden: Trotz der Länge der Sätze erzielen die Muttersprachler sehr hohe Werte bei der wörtlichen Satzwiedergabe. Möglicherweise wird der Vergleich

zwischen den beiden muttersprachlichen Gruppen durch Deckeneffekte erschwert und Material mit noch höherer Testschwierigkeit ist nötig, um die Unterschiede zwischen den unterschiedlich instruierten muttersprachlichen Gruppen aufzuzeigen. Auch ist es möglich, dass der numerische Unterschied zwischen den Gruppen bei einer größeren Stichprobe auch statistisch aufzeigbar wäre.

Problematisch ist der Vergleich zwischen den Studien 4 und 5 auch aufgrund des Wechsels der Modalität. In Studie 4 werden die Sätze am Bildschirm gelesen, in Studie 5 werden sie, wie in den vorangegangenen Studien, wieder auditiv präsentiert. Der Wechsel zur visuellen Präsentation in Studie 4 ist dem Aufbau der Lesestudie geschuldet, der sich nur schwerlich auf die auditive Präsentation übertragen lassen würde. Während der auditiven Präsentation ist damit aber auch nicht die Möglichkeit gegeben, das Tempo der Rezeption des Satzes den unterschiedlichen Anforderungen anzupassen. Dies schien nicht problematisch, da das stille Lesen eines Satzes gewöhnlich bei geübten Lesern weniger Zeit in Anspruch nimmt als das Sprechen des selben Satzes, die Verarbeitungszeit also beim Hören eines Satzes ausreichend lang auch für die Verarbeitung zum Behalten sein sollte. Veränderungen der grundsätzlichen Zusammenhänge wurden also durch den Modalitätswechsel nicht erwartet und sind aus den Vorannahmen nicht abzuleiten. Dennoch wird die zusammenhängende Interpretation durch diesen Modalitätswechsel zwischen den Experimenten erschwert.

Würden sich aber die beiden in den Studien 4 und 5 für die Muttersprachler gefundenen Effekte (die größere Belastung während des Leseprozesses zum wörtlichen Behalten im Vergleich zum inhaltlichen Behalten und das fast gleiche Abschneiden in der wörtlichen Wiedergabeleistung unter beiden Instruktionsbedingungen) in Folgeuntersuchungen bestätigen lassen, so würde dies die grundsätzliche Hypothese in Frage stellen. Diese besagt, dass Oberflächeninformation durch die kontrollierte Ausrichtung von Aktivierungsenergie aktiv gehalten wird. Bei der Instruktion zur inhaltlichen Wiedergabe sollte die Oberflächeninformation in geringerem Maße für die Reproduktion verfügbar sein als bei der Instruktion zur wörtlichen Wiedergabe. Während des Leseprozesses zum wörtlichen Behalten konnte in Studie 4 eine größere Belastung des Arbeitsgedächtnisses festgestellt werden (verglichen mit dem Leseprozess zum inhaltlichen Behalten). Lässt sich auch in weiteren Studien zeigen, dass diese erhöhte Belastung nicht zu einer besseren Leistung in der wörtlichen Wiedergabe führt als das weniger belastende Lesen zur inhaltlichen Wiedergabe, so muss daraus geschlossen

werden, dass zwar beim Lesen zum wörtlichen Behalten kontrollierte Prozesse für die Steuerung der Aktivierungsenergie sorgen, dies sich jedoch nicht auf die Wiedergabeleistung auswirkt. Diese Diskrepanz könnte einen einfachen Grund haben: Wie mehrfach betont wurde, ist die Aktivierungsenergie, ebenso wie die Aufmerksamkeit, eine begrenzte Ressource. Mehr Steuerung führt zu einem höheren kognitiven Aufwand und damit zu höherer Belastung - sie kann aber nicht mehr Aktivierungsenergie erzeugen. Wird diese schon für das Behalten zur inhaltlichen Wiedergabe ausgeschöpft (worauf die auch in dieser Bedingung nicht vollständige Wiedergabe hindeutet), so ist auch bei kontrolliertem Einsatz nicht mehr Aktivierungsenergie vorhanden - die Leistung verbessert sich also trotz des Einsatzes kontrollierter Prozesse nicht. Es besteht während des Lesens also eine höhere Belastung und kontrollierte Prozesse bemühen sich um die Aufrechterhaltung der Oberflächeninformation. Da sich die Aktivierungsenergie jedoch auf viele Informationen, die für die Reproduktion nötig sind, verteilt, sorgt auch die kontrollierte Zuweisung von Aktivierungsenergie an die Satzoberfläche nicht für eine signifikant höhere Wahrscheinlichkeit, diese zu erinnern, als es der weniger kontrollierte Fluss der Aktivierungsenergie beim Behalten zur inhaltlichen Wiedergabe tut. Diese Erklärung ist jedoch an diesem Punkt spekulativ. Auch besteht die Möglichkeit, dass die Unterschiede in der Rezeptionsphase, die durch die Instruktionen bedingt sind, stärker bei visueller selbstgesteuerter Präsentation als bei der auditiven Präsentation zum Tragen kommen, da bei Letzterer kein Einfluss auf das Präsentationstempo genommen werden kann. Bis Folgestudien diese Effekte bestätigt haben, sollte die ursprüngliche und aus der Forschung plausibel ableitbare Hypothese beibehalten werden.

Im nun folgenden Kapitel werden die Ergebnisse und Einschränkungen der für diese Arbeit durchgeführten Studien zusammengefasst und Implikationen für theoretische Modelle des Satzbehaltens in einer L2 diskutiert.

11 Allgemeine Diskussion

11.1 Annahmen und Ergebnisse im ersten Teil

Im ersten Teil dieser Arbeit (Studien 1 - 3) wurde die Annahme geprüft, dass die Leistung in der wörtlichen Satzwiedergabe die Sprachkompetenz von hochkompetenten Sprechern einer L2 (verglichen mit Aufgaben, die die allgemeine Sprachkompetenz messen und verglichen mit der Leistung von Muttersprachlern in diesen beiden Maßen) systematisch unterschätzt. In der wörtlichen Satzwiedergabe kann bei hochkompetenten Sprechern einer L2 ein aufgabenspezifisch schlechteres Abschneiden beobachtet werden. In Studie 1 wurde für Kinder im Vorschulalter (der derzeitigen Zielgruppe von Sprachstandserhebungen, die unter anderem diese Aufgabe verwenden) gezeigt, dass Kinder mit DaZ eine aufgabenspezifisch schlechtere Leistung in der wörtlichen Satzwiedergabe zeigen als aus ihrer Leistung in anderen Aufgabentypen und der Leistung von Kindern mit DaM in den verschiedenen Aufgabentypen zu erwarten wäre. In Studie 2 wurde gezeigt, dass dieser aufgabenspezifische Unterschied auch für erwachsene Sprecher einer L2 mit frühem Spracherwerb und muttersprachlicher Sprachkompetenz Bestand hat. Studie 3 zeigte ergänzend, dass der aufgabenspezifische Nachteil auf andere Sprachen übertragbar ist und auf die Verarbeitung einer L2 zurück zu führen ist, nicht auf die Bilingualität per se. Dies konnte in einer bilingualen Versuchsgruppe aufgezeigt werden, in der das aufgabenspezifisch schlechtere Abschneiden jeweils für die L2 gefunden wurde.

11.2 Annahmen und Ergebnisse im zweiten Teil

Im zweiten Teil dieser Arbeit sollte eine mögliche Alternative zur wörtlichen Satzwiedergabe, die inhaltliche Satzwiedergabe, auf ihre Tauglichkeit zur Messung der Sprachkompetenz bei *Near-Natives* hin überprüft werden. Dazu wurde im ersten Schritt die Annahme der geringeren Belastung des Arbeitsgedächtnisses durch die Anforderung zum inhaltlichen Behalten im Vergleich zum wörtlichen Behalten geprüft. Im zweiten Schritt wurde der Unterschied zwischen Sprechern einer L1 und einer L2 in beiden Aufgabenstellungen verglichen.

Die Annahme einer geringeren Belastung beim inhaltlichen im Vergleich zum wörtlichen Behalten konnte für Muttersprachler in Studie 4 bestätigt werden. Die Lesegeschwindigkeit der Teilnehmer, die instruiert wurden, die Sätze inhaltlich

wiederzugeben, war geringer als die Lesegeschwindigkeit der Teilnehmer, die instruiert wurden, die Sätze wörtlich wiederzugeben. Die niedrigere Lesegeschwindigkeit wurde als Indikator für eine höhere Belastung des Arbeitsgedächtnisses interpretiert: Durch die über den Satz hinweg stetig steigende Behaltensanforderung wird die Rezeption des folgenden Materials zum Behalten verlangsamt (vgl. Aaronson & Scarborough, 1976). Beim Lesen zum Behalten des Inhalts, aber nicht der Satzoberfläche, konnte eine höhere Lesegeschwindigkeit und eine geringere Abnahme dieser Geschwindigkeit über den Satz hinweg beobachtet werden als beim wörtlichen Behalten, was auf eine geringere Belastung des Arbeitsgedächtnisses hinweist.

Im zweiten Schritt wurde erwartet, dass diese geringere Behaltensanforderung im Vergleich zur wörtlichen Wiedergabe zu einer Annäherung der Leistung von Muttersprachlern und *Near-Natives* in ihrer Leistung in der inhaltlichen Satz wiedergabe führen würde: Bei vergleichbarer Sprachkompetenz sollten beide Gruppen auch in der inhaltlichen Satz wiedergabe ähnlich abschneiden, zumindest sollte der Unterschied zwischen den Gruppen geringer ausfallen als bei der wörtlichen Satz wiedergabe zu beobachten war. Diese Annahme konnte jedoch nicht bestätigt werden. Die Leistung in der inhaltlichen Satz wiedergabe bildet die Sprachkompetenz von *Near-Natives* scheinbar nicht besser ab als die wörtliche Satz wiedergabe. Die Interpretation der Unterschiede zwischen den *Near-Natives* und Muttersprachlern, die in Studie 5 instruiert wurden, die Sätze inhaltlich wiederzugeben, wird durch die sich widersprechenden Analysen (Prozentskala und z-Transformation) erschwert, insgesamt deuten diese aber darauf hin, dass der Unterschied zwischen den Sprachgruppen, der in der wörtlichen Satz wiedergabe gefunden wurde, auch in der inhaltlichen Satz wiedergabe Bestand hat. *Near-Natives* schneiden also bei vergleichbarer Sprachkompetenz auch in der inhaltlichen Satz wiedergabe schlechter ab als Muttersprachler. Darüber hinaus finden sich Indikatoren für Unterschiede zwischen den Teilnehmern mit L1 und L2 Deutsch in der Wirksamkeit der Instruktion zum wörtlichen Satz behalten. Alternativ können diese Unterschiede auch als Hinweise darauf gedeutet werden, dass die unterschiedliche Schwierigkeit für Sprecher einer L1 und L2 im wörtlichen Satz behalten sich auch auf die Leistung im inhaltlichen Satz behalten auswirkt: Die wörtliche Wiedergabe ist eine für Muttersprachler so einfache Aufgabe, dass auch bei der Möglichkeit zur inhaltlichen Wiedergabe wörtlich wiedergegeben wird. Dies führt bei den Muttersprachlern auch in der inhaltlichen Wiedergabe zu einem den *Near-Natives* überlegenen Leistungsniveau.

Auf Grundlage der Daten kann nicht entschieden werden, welche dieser möglichen Interpretationen zutreffend ist. Die Ergebnisse könnten auch ein Hinweis darauf sein, dass der Modalitätswechsel in der Präsentation zwischen den beiden Studien 4 und 5 kritisch zu betrachten ist. In jedem Fall liegt jedoch der Schluss nahe, dass die inhaltliche Wiedergabe von Sätzen keine sinnvolle Alternative zur wörtlichen Satzwiedergabe in Sprachtests darstellt. Der Aufgabentyp der Satzwiedergabe ist also auch mit dieser Modifikation nicht für das Testen in einer L2 geeignet.

11.3 Einschränkungen und Probleme der Studien 1, 2 und 3

Die einzelnen Studien im ersten Teil der Arbeit haben jede für sich genommen Schwächen, die durch die Arbeit in den Institutionen verursacht sind. Dennoch ergeben sie zusammengenommen ein klares Bild: In allen drei Studien zeigt sich, dass beim Testen in einer sehr gut beherrschten L2 die Leistung in der Satzwiedergabe stärker von der muttersprachlichen Leistung abweicht als in alternativen Maßen der Sprachkompetenz

Durch die nötige Feldforschung in Kindergärten und Schulen ergaben sich Einschränkungen, die sich nicht gänzlich abwenden ließen. In Studie 1 war der Faktor L1 (DaZ vs. DaM) mit dem Alter der Kinder konfundiert. Obwohl in mehreren Analysen gezeigt werden konnte, dass der gefundene Effekt sehr viel wahrscheinlicher auf die Sprachgruppe als auf das Alter der Kinder zurückgeführt werden kann, bleibt diese Konfundierung problematisch. Der jüngere Altersdurchschnitt der Kinder mit DaZ im Vergleich zu den Kindern mit DaM ist ein Zufall, der vielleicht durch eine größere Stichprobe mit einer (eventuell) gleichmäßigeren Altersverteilung hätte abgemildert werden können. Leider konnte ich in der für die Erhebung zur Verfügung stehenden Zeitspanne nicht mehr Kindergärten, Eltern und Kinder zur Teilnahme bewegen. Dies ist eventuell auf die hohe Belastung der Kinder mit Testverfahren in den Kindertagesstätten und das Misstrauen der Eltern gegenüber diesen Testverfahren zurück zu führen.

Studie 2 arbeitet mit einem Matchingverfahren. Problematisch ist hier im zweiten Teil der Vergleich zwischen *Near-Natives* und Muttersprachlern unterschiedlichen Alters. Der Vergleich mit jüngeren Muttersprachlern war jedoch notwendig, um eine muttersprachliche Vergleichsgruppe zu finden, die eine ähnlich niedrige Leistung in der Satzwiedergabe zeigt, wie die Gruppe der *Near-Natives*. Da die jungen Muttersprachler als

Gymnasiasten der erwachsenen Stichprobe der Studenten in vielen Punkten ähneln (mit der offensichtlichen Ausnahme ihres Alters), ist dieser Vergleich dennoch sinnvoll.

Studie 3 wurde am Deutsch-Französischen Gymnasium durchgeführt. Hier entstanden ungleich große Versuchsgruppen durch die nicht vorhersehbare höhere Teilnahmebereitschaft der deutschen Schüler im Vergleich zu den französischen Schülern. Da die Schüler in beiden Sprachen, Deutsch und Französisch, angesprochen wurden, ist nicht klar, worauf dieser Unterschied zurück zu führen ist. Leider war mit Rücksicht auf die schulischen Abläufe eine Nacherhebung auch in diesem Fall nicht möglich.

Für sich genommen weist jede einzelne Studie in Teil 1 dieser Arbeit Schwächen auf; insgesamt jedoch ergibt sich aus ihnen konvergierende Evidenz für einen aufgabenspezifischen Nachteil von Sprechern einer L2 in der wörtlichen Satz wiedergabe.

11.3.1 Diskussion der Vergleichsmaße der Sprachkompetenz: C-Test und SETK 3-5

In dieser Arbeit wird die Leistung von Zweitsprachlern in der Satz wiedergabe mit ihrer Leistung in einem alternativen Maß der Sprachkompetenz und den jeweiligen Leistungen von Muttersprachlern in den beiden Maßen in Beziehung gesetzt. Dem C-Test und dem SETK 3-5 kommen als den Vergleichsmaßen der Sprachkompetenz zentrale Rollen in dieser Arbeit zu. Im Folgenden soll noch einmal kurz diskutiert werden, warum diese Maße als valide Maße der Sprachkompetenz betrachtet werden können.

Für die Validität des C-Tests als Maß der allgemeinen Sprachkompetenz wurden weiter oben bereits zahlreiche Belege angeführt (einen weiteren guten Überblick über die Forschung zum C-Test gibt die Internetseite www.c-test.de, Raatz & Grotjahn, Stand 24.10.2011). Der C-Test basiert auf dem Prinzip der reduzierten Redundanz, er reduziert also die in der natürlichen Sprache vorhandene redundante Information. Beim Ergänzen der so entstandenen Lücken im C-Test werden sowohl das Verstehen als auch das Produzieren von Sprache getestet, sowie (durch die Beschränkung der Bearbeitungsdauer) die Automatisierung. Eine geringe Automatisierung würde zu einer längeren Bearbeitungsdauer des C-Tests und damit aufgrund des Zeitlimits zu einer geringeren Punktzahl führen. Der C-Test testet also prinzipiell die gleichen Fähigkeiten wie die Satz wiedergabe, jedoch ohne eine Behaltensanforderung zu stellen. Daher ist der C-Test in besonderer Weise als Vergleichsmaß der Sprachkompetenz für diese Arbeit geeignet. Der

C-Test ist außerdem ein ökonomisches Testverfahren, für das dennoch hohe Korrelationen mit ausführlicheren Tests des Sprachkompetenz aufgezeigt werden können (z.B. TestDaF, TOEFL).

Feine Unterschiede im Automatisierungsgrad zwischen Muttersprachlern und *Near-Natives* können durch den C-Test nicht erfasst werden. Jedoch ist dies nicht problematisch, da diese feinen Unterschiede ab einem hohen Automatisierungsgrad nicht mehr als Unterschiede in der Sprachkompetenz interpretiert werden sollten, da sie hinsichtlich der Kommunikationskompetenz in der muttersprachlichen Sprachgemeinde und im Hinblick auf die Chance auf eine erfolgreiche Bildungskarriere keine Aussagekraft besitzen. Die Teilhabe an Bildungschancen sicherzustellen ist (politisches) Ziel der Sprachstandserhebungen – für dieses ist der Automatisierungsunterschied irrelevant. Lediglich ein ausreichend hoher, kein muttersprachlicher Automatisierungsgrad ist hierfür relevant. Ein sehr hoher Automatisierungsgrad der Sprachverarbeitung ist Voraussetzung, um den C-Test in der vorgegebenen Zeit zu bearbeiten.

In der ersten Studie kam der SETK 3-5 (Grimm, 2001) als Vergleichsmaß der allgemeinen Sprachkompetenz zum Einsatz. Der SETK 3-5 ist ein etabliertes Werkzeug zur Bestimmung des Sprachentwicklungsstandes bei Muttersprachlern und die verwendeten Aufgabentypen sind (außer der Satzwiedergabe) grundsätzlich auch für den Einsatz mit Zweitsprachlern geeignet (vgl. Kany & Schöler, 2010). Die Bestimmung eines kombinierten Maßes aus den Aufgabentypen des SETK 3-5 (mit Ausschluss der Satzwiedergabe) ergab ein möglichst allgemeines Maß der Sprachkompetenz von Kindern mit DaZ und DaM. Wie erwartet schnitten die Kinder mit DaZ in allen Aufgabentypen durchschnittlich schlechter ab als die Kinder mit DaM. Dies ist darauf zurück zu führen, dass die Kinder mit DaZ noch keine ausreichende Anzahl an Kontaktmonaten mit der deutschen Sprache hatten, um zu den Kindern mit DaM aufzuschließen.

Beim Testen der Kinder mit DaZ und dem Vergleich ihrer Leistung mit muttersprachlichen Kindern sieht man sich zwei schwer zu vereinbarenden Zielstellungen gegenüber. Einmal soll die Sprachkompetenz der Kinder mit DaZ in einem Verfahren gemessen werden, das ihre Leistung valide abbildet und ihre besondere Situation, den Zweitspracherwerb, berücksichtigt. Gleichzeitig soll die gemessene Leistung mit der der Muttersprachler vergleichbar sein. Um diese beiden Ziele zu vereinbaren, wurde in dieser Arbeit auf einen Test zurückgegriffen (den SETK 3-5), der zwar für Muttersprachler

konzipiert wurde, jedoch flossen nur solche Aufgaben in die Bewertung der Sprachkompetenz mit ein, die auch für Zweitsprachler geeignet scheinen. Der Bedarf an Testverfahren, die speziell auf Kinder mit DaZ zugeschnitten sind, besteht jedoch für die Praxis weiter und wird durch diese Arbeit noch einmal betont.

11.4 Allgemeine Diskussion der Bedeutung der Ergebnisse für Theorie und Praxis

Automatisierungsunterschiede zwischen einer L1 und einer L2 existieren auch, wenn der Zweitspracherwerb im sehr frühen Kindesalter erfolgte und möglicherweise sogar beim simultanen Spracherwerb, wie Sebastián-Gallés et al. (2005) und Rinker et al. (2010) für die phonologische Verarbeitung zeigen konnten. Diese Automatisierungsunterschiede sind vermutlich für das schlechtere Abschneiden in der Satzwiedergabe in einer L2 (im Vergleich zur Leistung in anderen Aufgabentypen und der Leistung einer muttersprachlichen Vergleichsgruppe) verantwortlich. Dieser automatisierungsbedingte Verarbeitungsnachteil könnte jedoch durch einen Vorteil in der Leistung bei Prozessen, die kontrollierte Aufmerksamkeit benötigen, ausgeglichen oder moderiert werden. Wie bereits weiter oben diskutiert wurde, finden Luo et al. (2010) und auch Bialystok, Craik, Klein und Viswanathan (2004) Vorteile für bilingualen im Vergleich zu monolingualen Sprechern in Aufgaben, die neben der Speicherung von Inhalten auch deren Manipulation verlangen (z.B. Reihenfolgeinformationen, die rückwärts wiedergegeben werden müssen). Die Autoren führen die bessere Leistung der bilingualen Probanden in ihren Studien darauf zurück, dass diese geübt darin sind, irrelevante Informationen zu unterdrücken, da sie ständig zwei Sprachsysteme voneinander getrennt halten müssen und dabei Interferenz durch die gerade nicht gefragte Sprache verhindern müssen. Tatsächlich gibt es Hinweise darauf, dass bei der Verwendung einer L2 auch die Repräsentationen der L1 aktiviert werden (z.B. Preston & Lambert, 1969, Marian & Spivey, 2003, Hermans, Bongaerts, de Bot & Schreuder, 1998). Bialystok (2007) nimmt an, dass bilingualen Sprecher die gerade nicht verwendete Sprache unterdrücken, wodurch sie geübt darin sind, Interferenz zu verhindern. Weiterhin haben bilingualen Probanden Vorteile durch niedrigere Wechselkosten bei *task-switching*-Aufgaben und immer dann, wenn Information im Arbeitsgedächtnis nicht nur gehalten, sondern auch manipuliert werden muss. Überwiegend finden sich diese Vorteile jedoch in nicht-sprachbezogenen Aufgabenstellungen. Doch auch in diesen wird ein Vorteil für bilingualen Personen nur gefunden, wenn die Grenzen der Arbeitsgedächtniskapazität nicht überschritten werden (vgl. Bialystok, 2007). Da all diese Bedingungen nicht auf die Satzwiedergabe, wie sie in

dieser Arbeit und den Sprachstandserhebungen eingesetzt wird, zutreffen - sie behandelt sprachliches Material, das nicht manipuliert wird und überschreitet die Kapazität des Arbeitsgedächtnisses²⁴ - ist es sehr unwahrscheinlich, dass die Automatisierungsunterschiede in der L2 durch einen Vorteil in der effizienteren Arbeit der zentralen Exekutive ausgeglichen oder moderiert werden könnten. Ein weiterer Beleg gegen einen allgemeinen, also beide Sprachen betreffenden Nachteil für bilinguale Sprecher findet sich in den Ergebnissen aus Studie 3 der vorliegenden Arbeit: In einer bilingualen Stichprobe zeigt sich für hochkompetente Sprecher einer L2 ein aufgabenspezifischer Nachteil in der Satzwiedergabe in ihrer L2, auch wenn sie mit anderen Bilingualen verglichen werden, die diese L2 (Deutsch oder Französisch) jeweils als Muttersprache sprechen. Die Unterschiede finden sich also nicht zwischen bilingualen und monolingualen Sprechern, wie nach Studie 2 nicht auszuschließen war, sondern zwischen der Verarbeitung einer L1 und einer L2.

In Studie 5 konnte keine Verringerung des Leistungsunterschiedes zwischen Muttersprachlern und *Near-Natives* bei der Instruktion zur inhaltlichen Wiedergabe im Vergleich zur wörtlichen Wiedergabe der Sätze gefunden werden. Dies ist wahrscheinlich auf die nicht eindeutig interpretierbare Instruktion zur inhaltlichen Wiedergabe und die Metakognitionen, die diese Instruktion hauptsächlich bei den *Near-Natives* hervorruft, zurück zu führen. Daneben war auffällig, dass die Leistung in der wörtlichen Satzwiedergabe bei den Muttersprachlern nicht von der Instruktion abhängig war, während die kognitive Belastung in Studie 4 diese unterschiedlichen Anforderungen jedoch widerspiegelte (in Form geringerer Lesegeschwindigkeit beim wörtlichen im Vergleich zum inhaltlichen Behalten von Sätzen). Der Unterschied zwischen den beiden muttersprachlichen Gruppen in der wörtlichen Satzwiedergabe in Studie 5 ist nur numerisch, aber nicht statistisch in den Daten zu finden. Dies mag jedoch auf die geringe Stichprobengröße zurück zu führen sein. Auch könnten bei der hohen Leistung der Muttersprachler in der Satzwiedergabe bei einem Vergleich unter Muttersprachlern Deckeneffekte eine Rolle spielen, die die Unterschiede zwischen den beiden muttersprachlichen Gruppen abschwächen. Daneben ist der Modalitätswechsel in der Präsentation zwischen Studien 4 und 5 eventuell problematisch. Zeigt sich dieses Bild jedoch in weiteren Studien mit größerer Stichprobe und geringerer muttersprachlicher

²⁴ Die Sätze in dieser Arbeit und in den Sprachstandserhebungen werden in der Mehrzahl der Fälle nicht vollständig wiedergegeben. Dieser Verlust an Information deutet darauf hin, dass das Arbeitsgedächtnis in diesen Fällen über seine Grenzen hinaus beansprucht wird. In den Fällen, in denen die Sätze vollständig wiedergegeben werden, ist nicht von einer Überlastung des Arbeitsgedächtnisses, sondern lediglich von einer hohen Belastung auszugehen.

Wiedergabeleistung, so stellt es die grundsätzliche Hypothese in Frage, dass die Wiedergabe der Satzoberfläche kontrollierte Prozesse zur Wiedergabe beansprucht. Aus diesem einzelnen Ergebnis sollte dieser Schluss jedoch nicht vorschnell gezogen werden.

Das Ziel dieser Arbeit war es zu überprüfen, ob hochkompetente Sprecher einer L2 in ihrer Sprachkompetenz unterschätzt werden, wenn als dessen Maß ihre Leistung in der Satzwiedergabe herangezogen wird. Dazu wurde ihre Leistung in der Satzwiedergabe jeweils mit der in einem alternativen Maß der Sprachkompetenz und mit der von Muttersprachler verglichen. Die Ergebnisse aus beiden empirischen Teilen dieser Arbeit bestärken die Vorannahme, dass die Satzwiedergabe für hochkompetente Sprecher einer L2 kein valides Maß ihrer Sprachkompetenz darstellt, insbesondere, wenn ihre Leistung mit der von Muttersprachlern verglichen wird. Auch eine Modifikation der Aufgabe hin zum inhaltlichen statt wörtlichen Satzbehalten erhöht die Aussagekraft der Satzwiedergabe bezüglich der Sprachkompetenz von hochkompetenten Zweitsprachlern nicht. Dieser Befund hat eine einfache Implikation für die Praxis der Sprachstandserhebungen: Die Aufgabe "Sätze nachsprechen" sollte in diesen Tests nicht für Kinder mit DaZ zum Einsatz kommen. Die Implikationen für Modelle des Satzbehaltens werden im Folgenden dargelegt.

11.4.1 Implikationen für Modelle des Satzbehaltens in einer L2

Es konnte gezeigt werden, dass sich das Satzbehalten in einer L2 von dem in einer L1 unterscheidet. Beim Satzbehalten in der L2 steht weniger Aufmerksamkeit für die Aufrechterhaltung der Satzoberfläche zur Verfügung, als beim Satzbehalten in einer L1. Dieser Unterschied entsteht durch den unterschiedlichen Automatisierungsgrad der Sprachverarbeitung in einer L1 und L2. Letztere Annahme wurde nicht direkt getestet, kann aber aus der Forschungsliteratur zur Verarbeitung einer L2 bei hochkompetenten Sprechern und den hier präsentierten Daten plausibel abgeleitet werden. Die alternative Erklärung der Befunde durch einen Nachteil für bilinguale Sprecher insgesamt konnte ausgeschlossen werden. Der geringere Automatisierungsgrad bezieht sich demnach auf die Verarbeitung der L2 und schließt die Verarbeitung der L1 eines bilingualen Sprechers nicht mit ein.

Da die Prozesse des Verstehens bzw. Produzierens eines Satzes laufen bei der Satzwiedergabe überwiegend parallel ab: Bereits ab dem ersten Wort müssen Inhalt und

Struktur aktiv gehalten werden. Die Behaltensanforderung wächst bis zum Ende der Rezeption des Satzes und wird dann mit der beginnenden Reproduktion wieder geringer, da die bereits reproduzierten Inhalte nicht mehr behalten werden müssen. Werden Muttersprachler mit *Near-Natives* verglichen, so kann die Satzwiedergabe weder als Maß des verbalen Arbeitsgedächtnisses noch als Maß der Sprachverarbeitung betrachtet werden: Sie ist bestimmt von der Interaktion dieser beiden Prozesse, also der Sprachverarbeitung und des Behaltens. Da bei Muttersprachlern einer Altersstufe die Sprachverarbeitung vermutlich ähnlich automatisiert sein sollte, kann eine von der Vergleichsgruppe erheblich abweichende Leistung in der Satzwiedergabe als Defizit im verbalen Arbeitsgedächtnis gedeutet werden²⁵. Sie weist also zum Beispiel bei Kindern mit DaM auf eine Störung in der Sprachentwicklung hin, ist also valide in der Erfassung des Sprachstandes. Beim Testen mit Zweitsprachlern ist dieser Zusammenhang komplexer. Der Automatisierungsgrad der L2 weicht vermutlich vom muttersprachlichen Automatisierungsgrad ab. Diese Abweichung ist bei Kindern mit DaZ, die noch im Spracherwerbsprozess sind, größer als bei erwachsenen *Near-Natives*. Auf jeden Fall ist anzunehmen, dass sich der Automatisierungsgrad insbesondere während des L2-Erwerbs in der Kindheit verändert und dass sich auch erwachsene *Near-Natives* in ihrer Abweichung vom muttersprachlichen Automatisierungsgrad unterscheiden. Der Automatisierungsgrad von Zweitsprachlern kann also nicht exakt genug eingeschätzt werden, um den Einfluss der Sprachverarbeitung vom Einfluss des Behaltens der Information auf die gemessene Leistung zu trennen. Ein schlechtes Abschneiden in der Satzwiedergabe kann auf einen (für Zweitsprachler normalen) geringeren Automatisierungsgrad hindeuten, kann aber auch Anteile eines Defizits des verbalen Arbeitsgedächtnisses enthalten. Die Interpretation der Leistung von Zweitsprachlern in der Satzwiedergabe als Maß der allgemeinen Sprachkompetenz gestaltet sich also so komplex, dass sie nicht praktikabel ist.

Modelle der Satzwiedergabe in einer L2 sollten die Interaktion der Prozesse des Behaltens und der Sprachverarbeitung berücksichtigen. Die Ergebnisse deuten darauf hin, dass diese Prozesse, wie von Just und Carpenter (1992) beschrieben, auf eine gemeinsame Ressource zugreifen. Da diese Ressource beschränkt ist, wird eine aufwändigere, weil weniger automatisierte, Sprachverarbeitung sich negativ auf die für das Behalten verfügbare Ressource auswirken. Dieser Zusammenhang, der durch die bei Muttersprachlern überwiegend ähnliche Automatisierung der Sprachverarbeitung eine

²⁵ Alternativ würde eine schlechtere Leistung auf eine geringere Automatisierung der Sprachverarbeitung hindeuten, was jedoch für Muttersprachler (im Gegensatz zu Zweitsprachlern) ebenfalls auf einen gestörten Spracherwerb hindeuten würde.

geringere Rolle spielt, muss bei Modellen der Satzwiedergabe in einer L2 stärker berücksichtigt werden.

11.5 Weiterführende Forschungsfragen

Diese Arbeit hatte einerseits zum Ziel, den grundlegenden Effekt des aufgabenspezifisch schlechteren Abschneidens von hochkompetenten Sprechern einer L2 in der wörtlichen Satzwiedergabe aufzuzeigen. Zum anderen sollte eine mögliche Alternative zu dieser Aufgabe untersucht werden. Das Erreichen des ersten dieser beiden Ziele und das Verfehlen des zweiten werfen jedoch zahlreiche Fragen auf, die mit dem Effekt und den Studien im Zusammenhang stehen. Zentrale Punkte sind dabei die Untersuchung der Automatisierungsunterschiede, die Variation der alternativen Maße der Sprachkompetenz und der Einfluss der L1 auf die Leistung in der L2. Diese Punkte sollen in den folgenden Abschnitten kurz dargelegt werden.

11.5.1 Untersuchung der Verarbeitungsunterschiede

In Studie 4 konnte für Muttersprachler ein Unterschied in der Lesegeschwindigkeit von Sätzen aufgezeigt werden, je nachdem ob diese später wörtlich oder inhaltlich wiedergegeben werden sollten. Dieser wurde als geringere Belastung des Arbeitsgedächtnisses beim Lesen zum inhaltlichen Behalten im Vergleich zum wörtlichen gedeutet. Da sich die geringere Belastung des Arbeitsgedächtnisses bei den Zweitsprachlern in Studie 5 nicht in besserer Leistung in der Satzwiedergabe niederschlug, wäre es ein wichtiger nächster Schritt, die Enkodierungsunterschiede, die in Studie 4 für die Muttersprachler gefunden wurden auch für *Near-Natives* aufzuzeigen. Damit könnte überprüft werden, ob sich Muttersprachler und *Near-Natives* bereits in dieser Phase des Satzbehaltens und in Abhängigkeit von der Instruktion zum wörtlichen oder inhaltlichen Behalten unterscheiden. Dazu würde der Aufbau von Studie 4 mit *Near-Natives* wiederholt werden.

Die Automatisierungsunterschiede zwischen einer L1 und L2 sollten ebenfalls genau untersucht werden. Obwohl die theoretischen Annahmen Automatisierungsunterschiede ins Zentrum rücken, werden diese in den vorliegenden Studien nicht direkt erfasst. Dies liegt zum Teil daran, dass diese Unterschiede überwiegend durch bildgebende

Verfahren aufgezeigt werden. Der Einsatz dieser Verfahren lag nicht im Rahmen der Möglichkeiten dieser Studien, während derer in weiten Teilen Kinder in ihren Tageseinrichtungen oder Schüler direkt in der Schule getestet wurden. In einem nächsten Schritt sollten diese Zielgruppen (und andere) unter kontrollierten Laborbedingungen betrachtet werden.

11.5.2 Variation im C-Test-Verfahren: erhöhter Zeitdruck

Wie oben ausgeführt wurde, erfasst der C-Test durch die zeitliche Begrenzung der Bearbeitungsdauer teilweise die Automatisierung von Sprache. Jedoch war die zeitliche Begrenzung (die laut Vorgabe des C-Test-Verfahrens gewählt wurde) so bemessen, dass für Muttersprachler und *Near-Natives* (knapp) ausreichend Zeit zur Komplettierung des Tests zur Verfügung stand. Es wäre in einer folgenden Studie interessant, die Bearbeitungsdauer des C-Tests zu messen oder diese so zu begrenzen, dass auch sprachlich kompetente Teilnehmer den Test nicht vollständig ausfüllen können. Ein Vergleich zwischen Bearbeitungsdauer bzw. Leistung in einem zeitlich stärker begrenzten C-Test-Verfahren sollte Rückschlüsse auf Unterschiede in der Automatisierung der Sprachverarbeitung ermöglichen. Im nächsten Schritt könnte dann untersucht werden, ob sich Muttersprachler und *Near-Natives*, die in einem solchen Verfahren keine Unterschiede zeigen, dennoch in der Satzwiedergabe unterscheiden. Dies wäre ein weiterer starker Hinweis auf die Unangemessenheit des Testens mit der Aufgabe Satzwiedergabe in einer L2 und ebenso auf die Verschiedenheit der Prozesse zwischen einer L1 und L2 beim Satzbehalten²⁶.

11.5.3 Einsatz alternativer vergleichender Maße der Sprachkompetenz

Weiterführende Forschung sollte sich auch mit alternativen Vergleichsmaßen der Sprachkompetenz auseinandersetzen. Als Maße der Sprachkompetenz wurden bei den Kindern der SETK 3-5 und bei den Schülern und Erwachsenen deutsche und teilweise französische Tests nach dem C-Test-Prinzip eingesetzt. Obgleich beide Tests, wie dargelegt wurde, gute Maße der Sprachkompetenz sind, ist es nötig, den hier aufgezeigten Effekt auch anhand anderer Sprachtests zu replizieren. Für das Testen im Kindergartenalter

²⁶ Dabei sollten die Stichproben aus beiden Sprachgruppen groß genug sein, um Varianz, die z.B. durch Unterschiede im Schreibtempo entstehen kann, auszugleichen.

könnte dazu zum Beispiel der Heidelberger Sprachentwicklungstest (HSET, Grimm & Schöler, 1991) eingesetzt werden. Da zahlreiche Testverfahren für Kinder im Kindergartenalter zur Verfügung stehen, sollte die Replikation mit Kindern nicht problematisch sein. Dies gilt jedoch nicht für ältere Schüler und Erwachsene mit L2 Deutsch. Für diese Altersgruppen konnte ich neben dem C-Test keinen Test finden, die valide Aussagen über die Leistung in einer L1 und in einer L2 für Sprecher mit einem nahezu muttersprachlichen Kompetenzgrad machen kann. Der Einsatz des TestDaFs in seiner höchsten Schwierigkeitsstufe ist wahrscheinlich noch immer durch Deckeneffekte begrenzt. Daneben ist der TestDaF mit mehr als drei Stunden Durchführungsdauer ein sehr aufwändiges Testverfahren. Auch ist seine Eignung für den Einsatz mit Muttersprachlern nicht gegeben. Letzteres wäre aber nötig, um ein Maß zu haben, anhand dessen sich die Leistung von Muttersprachlern und Zweitsprachlern vergleichen ließe. Alternativ könnten Tests der verbalen Intelligenz zum Einsatz kommen, sofern die verwendeten Aufgaben für die Messung in einer L2 geeignet erscheinen, also gemessene Unterschiede nicht auf Automatisierungsunterschiede zurückzuführen sind.

11.5.4 Modalität in Studie 5

Zwischen den Studien 4 und 5 wurde die Modalität der Darbietung verändert. Um zu überprüfen, welchen Effekt die Darbietungsmodalität auf die Ergebnisse in der L1 und L2 und vor allem auf die Instruktion zum wörtlichen oder inhaltlichen Behalten von Sätzen hat, sollte Studie 5 repliziert werden unter der Verwendung der selbstgesteuerten visuellen Präsentation aus Studie 4. Eine größere Kontrolle über die Geschwindigkeit der Präsentation könnte zu einer je nach Instruktion stärker abweichenden Wiedergabeleistung führen. Die Abhängigkeit der (wörtlichen und inhaltlichen) Wiedergabeleistung von der Instruktion sollte auch zwischen den Sprachgruppen noch einmal unter diesen Bedingungen verglichen werden.

11.5.5 Einfluss der L1

Die Gruppen der *Near-Natives* in den Studien 2 und 5, sowie die Gruppe der Kinder mit DaZ waren bezüglich ihrer L1 heterogen. In diesen Studien wurde nicht berücksichtigt, welchen Einfluss die L1 der *Near-Natives* auf ihre Leistung in der L2 hat. Es wäre möglich, dass große Überschneidungen in der Syntax oder Phonologie zweier

Sprachen die Satzwiedergabe in der L2 erleichtern. Einige Modelle der bilingualen Sprachverarbeitung gehen von zumindest teilweise überlappenden Repräsentationen aus (z.B. Hartsuiker, Pickering & Veltkamp, 2004; Kroll & Stewart, 1994). Die Verarbeitung einer L2 könnte immer dann einen höheren Automatisierungsgrad erreichen, wenn ihre Prozesse mit denen der L1 überlappen. Umgekehrt könnte eine größere Ähnlichkeit der beiden Sprachen auch zu größerer Interferenz führen. Die möglichen Interaktionen der L1 und L2 in der Satzwiedergabe könnten ein Thema weitergehender Forschung sein.

Diese und weitere Aspekte des aufgezeigten grundlegenden Effekts verdienen Beachtung. Die Basis dieser weiterführenden Fragen bildet der Befund, dass Sprecher einer L2 in der wörtlichen Satzwiedergabe eine schlechtere Leistung zeigen, als anhand ihrer Leistung in alternativen Testverfahren und der Leistung einer muttersprachlichen Vergleichsgruppe zu erwarten ist. Nachdem die theoretischen Schlussfolgerungen, die sich daraus ergeben, bereits erläutert wurden, soll im nun folgenden Schlusswort in aller Kürze auf die Implikationen für die Praxis der Sprachstandserhebungen eingegangen werden.

12 Schlusswort

Die hier dargestellten Befunde unterstreichen den Bedarf an neuen Verfahren zur Erhebung des Sprachstands für Kinder mit DaZ. Die Aufgabe "Sätze nachsprechen" ist nicht als Teilaufgabe zur Messung der Sprachkompetenz von Kindern mit DaZ geeignet. Auch andere Aufgabentypen, die sich für Kinder mit DaM bewährt haben, sollten vor dem Einsatz mit DaZ-Kindern einer kritischen Prüfung unterzogen werden. Daneben sollten auch Erhebungen, die zur Evaluation von Sprachfördermaßnahmen durchgeführt werden, auf ihre Tauglichkeit hin überprüft werden. Beim Einsatz der Aufgabe "Sätze nachsprechen" würde sich zum Beispiel auch eine vorhandene Verbesserung der Sprachkompetenz bei Kindern mit DaZ nicht zeigen, da die Leistung der Kinder mit DaZ in diesem Aufgabentyp wenig Bezug zu ihrer Sprachkompetenz hat.

Sprachstandserhebungen in Kindergärten und -tagesstätten, die auch für Kinder mit DaZ eingesetzt werden, sollten auf ihre Validität für diese Zielgruppe hin geprüft werden. Nicht nur fehlende Normierungen und die Vernachlässigung der Kontaktmonate mit der Zweitsprache, sondern auch die Aufgabenstellungen per se könnten sie für den Gebrauch mit Kindern mit DaZ ungeeignet machen. Dies konnte hier bereits für die Aufgabe "Sätze nachsprechen" gezeigt werden.

Literatur

- Aaronson, D. & Scarborough, H.S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology*, 2 (1), 56-70
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Arras, U., Eckes, T., & Grotjahn, R. (2002). C-Tests im Rahmen des ‚Test Deutsch als Fremdsprache‘ (TestDaF): Erste Forschungsergebnisse. In: R. Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen: Bd. 4*. (S. 175-209). Bochum: AKS-Verlag.
- Baddeley, A.D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158-173.
- Baddeley, A.D. & Hitch, G.J. (1974). Working Memory. In G.A. Bower (Hrsg.) *The Psychology of Learning and Motivation* (S. 47–89) Academic Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 14, 575-589.
- Baddeley, A.D. (1986). *Working Memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4 (11), 417–422.
- Baur, R. S. & Meder, G. (1994). C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und in der Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In R. Grotjahn (Hrsg.) *Der C-Test. Theoretische Grundlagen und praktische Anwendungen: Band 2*, (S. 151-178), Bochum: Brockmeyer.
- Berth, F. (2008). Lernziel Deutsch, *Süddeutsche Zeitung*, 23.06.2008, München.
- Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism*, 10 (3), 210-223.
- Bialystok, E. (1997). The structure of age: in search of barriers to second language acquisition. *Second Language Research*, 13 (2), 116-137.
- Bialystok, E., Craik, F. I. M., Klein, R. & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence From the Simon Task. *Psychology and Aging*, 19(2), 290-303.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (Hrsg.), *Handbook of bilingualism: Psycholinguistic approaches* (S. 109–127). New York: Oxford University Press.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56 (S1), 9-49.

- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Hrsg.), *Linguistic perspectives on second language acquisition* (S. 41–68). Cambridge, UK: Cambridge University Press.
- Bosch, L., Costa, A. & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12 (2), 189-221.
- Brewer, W.F., Sampaio, C. & Barlow, M.R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language*, 52, 618-627.
- Bundesamt für Migration und Flüchtlinge (Hrsg.) (2010). Migrationsbericht 2008. Berlin.
- Bußmann, H. (1990). *Lexikon der Sprachwissenschaft*. Alfred Kröner Verlag: Stuttgart.
- Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (Hrsg.), *Language production Vol. 1: Speech and talk* (S. 423-459). London: Academic Press.
- Caplan, D. & Waters G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral Brain Science*, 22, 114 –126.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.
- CDU & FDP (Hrsg.) (2005). *Koalitionsvereinbarung zwischen CDU und FDP zur Bildung einer neuen Landesregierung in Nordrhein-Westfalen*. Düsseldorf
- CDU, CSU & FDP (Hrsg.) (2009). *Wachstum. Bildung. Zusammenhalt. Der Koalitionsvertrag zwischen CDU, CSU und FDP*. Berlin
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: Massachusatts.
- Clahsen, H. & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10 (12), 564-570.
- Conrad, R. & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, 55, 429-432.
- Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Contraints Within the Human Information Processing System. *Psychological Bulletin*, 104, 163 – 191
- Cowan, N. (1993). Activation, attention and short-term memory. *Memory and Cognition*, 21 (2), 162-167.
- Cowan, N. (1994). Mechanisms of verbal short-term memory. *Current Directions in Psychological Science*, 3 (6), 185-189.

- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Hrsg.) *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.
- Cowan, N., Sauls, J. S., & Brown, G. D. A. (2004). On the auditory modality superiority effect in serial recall: Separating input and output factors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 639-644.
- Cowan, N. & Morey, C.C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science*, 18 (8), 686-688.
- Crocker, M.W. & Pickering, M. (2000). *Architectures and mechanisms for language processing*. New York: Cambridge University Press.
- Cutler, A. (1988). The perfect speech error. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin*. London: Routledge.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450-466.
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Multilingual Matters: Bristol.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533.
- Dell, G. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, 93, 283-321.
- Dell, G. & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611-629.
- Dell, G., Chang, F. & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, 23(4), 517-542.
- Dell, G.S., Juliano, C., & Govindjee, A. (1993). Structure and Content in Language Production: A Theory of Frame Constraints in Phonological Speech Errors. *Cognitive Science*, 17, 149-195.
- Desmette D., Hupet M., Schelstraete M. A., & Van der Linden, M. (1995). Adaptation en langue française du « Reading Span Test » de Daneman et Carpenter (1980). *L'année psychologique*, 95(3), 459-482.
- Diller, J. & Jordan-Diller, K. (2003). Sentence repetition testing (SRT) and language shift survey of the Tuki language. *SIL Electronic Survey Reports*, 10.
- Eckes, T. (2010). Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In R. Grotjahn (Hrsg.), *Der C-Test: Beiträge aus der aktuellen Forschung*, (S. 125-192). Frankfurt: Lang.

- Engelkamp, J. & Rummer, R. (1999). Die Architektur des mentalen Lexikons. In A.D. Friederici (Hrsg.), *Enzyklopädie der Psychologie. Sprachrezeption C III, Sprache 2*. (S. 155-193) Göttingen: Hogrefe.
- Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (S. 102-134). New York: Cambridge University Press.
- Freneck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Herrida and J. Altarriba (Eds.) *Bilingual Sentence Processing*, (S. 217-236). North Holland.
- Fried, L. (2007). *Delfin 4 - Diagnostik, Elternarbeit und Förderung der Sprachkompetenz Vierjähriger in NRW. Screening Stufe I*. Technische Universität Dortmund.
- Fried, L. (2008). *Delfin 4 - Diagnostik, Elternarbeit und Förderung der Sprachkompetenz Vierjähriger in NRW. Screening Stufe I*. Technische Universität Dortmund.
- Fried, L., Briedigkeit, E., Isele, P. & Schunder, R. (2009). Delfin 4 - Sprachkompetenzmodell und Messgüte eines Instrumentariums zur Diagnose, Förderung und Elternarbeit in Bezug auf die Sprachkompetenz vierjähriger Kinder. *Zeitschrift für Grundschulforschung*, 2, 13-26.
- Fried, L., Briedigkeit, E. & Schunder, R. (2009). *Delfin 4 - Sprachförderorientierungen. Eine Handreichung*. Düsseldorf.
- Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6 (2), 78-84.
- Fromkin, V.A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47 (1), 27-52.
- Garrett, M.F. (1976). Syntactic processes in sentence production. In R. Wales & E. Walker (Hrsg.), *New approaches to language mechanisms* (pp. 231-255). Amsterdam: North-Holland.
- Garrett, M. F. (1984). The organization of processing structure for language production: Application to aphasic speech. In D. Caplan, A. R. Lecours, & A. Smith (Hrsg.) *Biological perspectives on language* (S. 172–193). Cambridge, MA: MIT Press.
- Grimm, H. (2001). Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5). Göttingen Hogrefe.
- Grimm, H. & Schöler, H. (1978/1991). Heidelberger Sprachentwicklungstest (HSET). Göttingen Hogrefe.
- Grotjahn, R. (1992). Der C-Test im Französischen. Quantitative Analysen. In R. Grotjahn (Hrsg.) *Der C-Test. Theoretische Grundlagen und praktische Anwendungen (Band 1)*. (S. 205-255). Bochum: Brockmeyer.

Grotjahn, R. (2007). The C-Test bibliography. Ruhr-Universität Bochum. <http://www.c-test.de/deutsch/index.php?lang=de&content=bibliografie§ion=c-test> (09.05.2011)

Guillelmon, D. & Grosjean, F. (2001). The gender marking effect in spoken word recognition. The case of bilinguals. *Memory and Cognition*, 29 (3), 503-511.

Hacker, W., Handrick, S. & Veres, T. (1996). Lesespannentest. Manuskript. Universität Dresden.

Hartsuiker, R.J., Pickering, J. & Veltkamp, E. (2004). Is Syntax Separate or Shared Between Languages? Cross-Linguistic Syntactic Priming in Spanish-English Bilinguals. *Psychological Science*, 15 (6), 409-414.

Hartsuiker, R.J., Kolk, H.H.J. & Huiskamp, P. (1999). Priming Word Order in Sentence Production. *The Quarterly Journal of Experimental Psychology*, 52A (1), 129-147.

Hermans, D., Bongaerts, T., de Bot, K. & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1, 213-229.

Jefferies, E., Lambon Ralph, M.A. & Baddeley, A.D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, 51 (4), 623-643.

Jescheniak, J.D., & Levelt, W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824-843.

Johnson, J. & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 20, 60-99.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.

Kany, W. & Schöler, H. (2010). *Fokus: Sprachdiagnostik: Leitfaden zur Sprachstandsbestimmung im Kindergarten (2. Auflage)*. Cornelsen Scriptor: Berlin.

Kintsch, W. & van Dijk, T.A.(1978). Toward a Model of Text Comprehension and Production. *Psychological Review*, 85 (5), 363 - 394

Kroll, J.F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representation. *Journal of Memory and Language*, 33, 149-174.

Kutas, M. & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4 (12), 463-470.

Kutas, M. & Hillyard, S.A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition* 11, 539-550.

- Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203-205.
- Levelt, W.J.M. (1993). Timing in speech production: With special reference to word form encoding. *Annals of the New York Academy of Sciences*, 682, 283-295.
- Levelt, W.J.M., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 615-618.
- Levelt, W.J.M., Roelofs, A. & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-38.
- Lombardi, L. & Potter, M.C. (1992). The Regeneration of Syntax in Short Term Memory. *Journal of Memory and Language*, 31, 713-733.
- Luo, L., Luk, G. & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, 114 (1), 29-41.
- Majerus, S. (2010). Das verbale Kurzzeitgedächtnis als Produkt der Interaktion zwischen Aufmerksamkeitskapazitäten, Sequenzverarbeitung und Aktivierung des Sprachsystems. *Psychologische Rundschau*, 61 (1), 10-17.
- Marian, V. & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6 (2), 97-115.
- McLaughlin, B. (1978). *Second Language Acquisition in Childhood*. Hillsdale, N.J.: Erlbaum.
- Meyers, J.E., Volkert, K. & Diep, A. (2000). Sentence Repetition Test: Updated norms and clinical utility. *Applied Neuropsychology*, 7 (3), 154-159.
- Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen & Ministerium für Generationen, Familie, Frauen und Integration des Landes Nordrhein-Westfalen (Hrsg.) (2009). *Feststellung des Sprachstands zwei Jahre vor der Einschulung, Fachinformation zum Verfahren 2009*. Düsseldorf.
- Moors, A. & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, 132 (2), 297-326.
- Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78, 679-684.
- Navarra, J., Sebastián-Gallés, N. & Soto-Faraco, S. (2005). The Perception of Second Language Sounds in Early Bilinguals: New Evidence From an Implicit Measure. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 912-918.
- Neumann, O. (1992). *Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen*. Psychologische Rundschau, 43, 83-101.

- Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology*, 50, 257-269.
- Oberauer, K. & Kliegl, R. (2010). Interferenz im Arbeitsgedächtnis: Ein formales Modell. *Psychologische Rundschau*, 61 (1), 33-42.
- Osterhout, L. & Holcomb, P.J. (1992). Event-Related Brain Potentials Elicited by Syntactic Anomaly. *Journal of Memory and Language* 31, 785-806.
- Park, N. & Martin, R. C. (2002). *A contribution of phonological representations to immediate sentence recall*. Poster presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York City.
- Perani, D., Abutalebi, J., Paulescu, E., Brambati, S., Scifo, P., Cappa, S.F. & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19 (3), 170-182.
- Potter, M.C. & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29, 633-654.
- Potter, M.C. & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265-282.
- Preston, M.S. & Lambert, W.E. (1969). Interlingual interference in a bilingual version of the stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, 8 (2), 295-301.
- Raatz, U. & Klein-Braley, C. (1982). The C-test – a modification of the cloze procedure. In T. Culhane, C. Klein-Braley & D. K. Stevenson (Hrsg.), *Practice and problems in language testing IV* (pp. 113-138). Colchester: University of Essex, Dept. of Language and Linguistics.
- Raatz, U. & Grotjahn, G. (2011) *C-Test - Der Sprachtest*. www.c-test.de (24.10.2011).
- Radloff, C.F. (1991). *Sentence repetition testing for studies of community bilingualism*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics.
- Reich, H.H. & Roth, H. J. (2004). Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger - HAVAS 5. Landesinstitut für Lehrerbildung und Schulentwicklung Hamburg.
- Rinker, T., Alku, P., Brosch, S. & Kiefer, M. (2010). Discrimination of native and non-native vowel contrasts in bilingual Turkish–German and monolingual German children: Insight from the Mismatch Negativity ERP component. *Brain & Language* 113, 90–95.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107-142.
- Rumelhart, D.E. & McClelland, J.L. (Hrsg.) (1986) *Parallel Distributed Processing*. Cambridge, London: MIT Press.

- Rummer, R. & Fiebach, C. (2010). Prozessbasierte Ansätze in der aktuellen Arbeitsgedächtnisforschung - Editorial. *Psychologische Rundschau*, 61 (1), 1-2.
- Rummer, R. (2003). Das kurzfristige Behalten von Sätzen. *Psychologische Rundschau*, 54 (2), 93-102.
- Rummer, R. & Engelkamp, J. (2011). Towards a language based account on verbal working memory. In C. Zelinsky-Wibbelt (Hrsg.), *Relations between Language and Memory: Organization, Representation, and Processing*. Lang, Peter Frankfurt.
- Rummer, R. & Schweppe, J. (2009). *Nachsprechen von Sätzen und Kunstwörtern: Implikationen für die Sprachstandsdiagnose von Kindern mit Migrationshintergrund*. Antrag auf Gewährung einer Sachbeihilfe bei der Deutschen Forschungsgemeinschaft (DFG).
- Rummer, R. & Schweppe, J. (2005). Evidence for a modality effect in sentence retention. *Psychonomic Bulletin & Review*, 12, 1094-1099.
- Rummer, R. & Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language*, 45, 451-467.
- Rummer, R. & Engelkamp, J. (2003). Phonological information in immediate and delayed sentence recall. *The Quarterly Journal of Experimental Psychology*, 56A, 83-95.
- Rummer, R., Engelkamp, J. & Konieczny, L. (2003). *The subordination effect: Evidence from self-paced reading and recall*. The European Journal of Cognitive Psychology, 15 (4), 539-566.
- Rummer, R. (1996). *Kognitive Beanspruchung beim Sprechen*. Weinheim: PVU.
- Salamé, P., & Baddeley, A.D. (1982). Disruption of memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 150-164.
- Schmitterer, A., Knackstedt, D., Salomon, V., Scholze, V., & Schweppe, J. (2009). *Systematischer Nachteil von Zweitsprachlern bei der Satzwiedergabe*. Unveröffentlichte Daten, Universität Erfurt.
- Schneider, W. & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G.H. Bower (Hrsg.) *The psychology of learning and motivation: advances in research and theory*. (S. 54-121) San Diego: Academic Press.
- Schulz, P. & Tracy, R. (in Druck). *LiSeDaZ - Linguistische Sprachstandserhebung - Deutsch als Zweitsprache*. Göttingen: Hogrefe.
- Schulz, P., Tracy, R. & Wenzel, R. (2008). Entwicklung eines Instruments zur Sprachstandsdiagnose von Kindern mit Deutsch als Zweitsprache: Theoretische Grundlagen und erste Ergebnisse. In: B. Ahrenholz (Hrsg.), *Zweitspracherwerb: Diagnosen, Verläufe, Voraussetzungen*. (S. 9-33), Freiburg: Fillibach.

Schweppe, J. (2006). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. Saarbrücken: Dissertation an der Universität des Saarlandes.

Schweppe, J. & Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory and Language*, 56, 336-356.

Schweppe, J., Rummer, R. & Fürstenberg, A. (2009). Beyond sentence boundaries: Grammatical gender information in short-term recall of texts. *Memory & Cognition*, 37, 73-80.

Sebastián-Gallés, N. & Bosch, L. (2005). Phonology and Bilingualism. In J.F. Kroll & A.M.B. De Groot (Hrsg.) *Handbook of Bilingualism: Psycholinguistic Approaches*. (S. 68-84) Oxford University Press.

Sebastián-Gallés, N., Echeverría, S. & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52 (2), 240-255.

Segalowitz, N. & Hulstijn, J. (2005). Automaticity in Bilingualism and Second Language Learning. In J.F. Kroll & A.M.B. De Groot (Hrsg.) *Handbook of Bilingualism: Psycholinguistic Approaches*. (S. 371-386) Oxford University Press.

Shiffrin, R. M. (1993). Short-term memory: A brief commentary. *Memory & Cognition*, 21, 193-197.

Spiewak, M. (2010). Zu kurz, zu spät, zu abstrakt. *Die Zeit*, 22.10.2010.

Spivey, M.J. & Tanenhaus, M.K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1521-1543.

Stowe, L.A. & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, 43 (4), 329-353.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268 (5217), 1632-1634.

Vigliocco, G. & Hartsuiker, R.J. (2002). The Interplay of Meaning, Sound, and Syntax in Sentence Production. *Psychological Bulletin*, 128 (3), 442-472.

Weber-Fox, C.M. & Neville, H.J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neurosciences*, 8, 231-256.

Weinert, S. (2006). Sprachentwicklung einschließlich Bilingualismus. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 656-664). Göttingen: Hogrefe.

Werker, J.F. & Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior Development*, 7, 49-63.

Werker, J.F. & Tees, R.C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 509-510.

Westfälische Rundschau (o.V.). *Bei Sprachförderung Ziel verfehlt*. 20.01.2010.

Wockenfuß, V. & Raatz, U. (2006). Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In R. Grotjahn (Hrsg.) *Der C-Test: Theorie, Empirie, Anwendungen. The C-Test: Theory, Empirical Research, Applications*. (S. 211-242) Frankfurt am Main: Lang.

Yip, V. & Matthews, S. (2007). *The bilingual child: early development and language contact*. Cambridge. Cambridge University Press.

Anhang

A Materialien

A1. Adaptierte Sätze aus dem Lesespannentest (Anzahl der Worte in Klammern)

TEST 1 Ich saß auf dem Stuhl und leise kam fremdländische Musik aus dem Radio der verschlossenen Küche nebenan. (17)

TEST 2 Dieser innige Austausch kann sich häufig auch dort ergeben, wo beide Teilnehmer einen ganz verschiedenen Beruf ausüben. (17)

TEST 3 Diese Tiere haben sich, wie jeden Winter, unter der Schneedecke Gänge angelegt und Nester aus Gras gebaut. (17)

TEST 4 Wer im Herbst fleißig Eicheln und Kastanien sammelt, schafft für den Winter geeignete Nahrung für die Tiere. (17)

TEST 5 Bei einem kurzen Besuch der Messe unterstrichen die Gäste die große Bedeutung der Produkte der Sportindustrie. (16)

TEST 6 Das Grau der Häuser um mich herum wirkte in diesem Moment so traurig auf mich wie noch nie zuvor. (19)

TEST 7 Zornig machte er mit der Hand eine befehlende Geste, und sie ging zögernd durch die offene Tür. (17)

TEST 8 Und dort hatte Uwe dann die glorreiche Idee, in Zukunft Boxen als Sport betreiben zu wollen. (16)

1. Aber so gelang es wohl niemandem, ihn vom Ofen weg und aus der inzwischen sehr verdreckten Hütte zu bringen. (19)

2. Wenn solch ein wichtiges Problem nicht richtig geklärt ist, muss man es im Kollegium ansprechen und diskutieren. (17)

3. Für Anna war die Woche Strandurlaub keine Erholung gewesen, ständig kreisten ihre Gedanken um den Sohn der Vermieterin. (18)

4. An einem sonnigen Abend im August waren Rainer und seine Schwester mit Moped und Rucksack in Halle unterwegs. (18)

5. Und auch seine Prinzipien übertrug er konsequent auf die Schüler und die ganze Schule, die er seit August leitete. (19)

6. Eva setzte sich auf den Stuhl vom Herrn Direktor, machte ein ganz ernstes Gesicht und Klaus musste lachen. (18)

7. Er fuhr mit dem Moped ohne Helm viel zu schnell auf der nassen Straße und kam ins Rutschen. (18)

8. Wenn irgend möglich, setze ich mich in die Straßenbahn, fahre ich nach außerhalb und gehe ein bisschen im Wald spazieren. (20)
9. Er sah weiter das halb helle, halb dunkle Gesicht seines Vaters an und erwartete den unvermeidlichen Wutanfall. (17)
10. Die Stadt verfügt seit mehreren Jahren über ein modernes Krankenhaus mit rund 1000 Betten und bester technischer Ausrüstung. (18)
11. Da das Konzert im Freien stattfand, musste es wegen des einsetzenden Sturms abgebrochen und die Zuschauer in Sicherheit gebracht werden. (20)
12. Er war seit dem Mittagessen nicht ins Haus zurückgegangen, um ein Glas Milch mit Honig zu trinken. (17)
13. Er fertigte sein letztes Werkstück kurz vor Feierabend ohne jegliches Interesse an und gab es gleichgültig zur Kontrolle. (18)
14. In pechschwarzer Nacht stand Elke in ihren ältesten Mantel gehüllt an der Kreuzung unter einem Baum im Regen. (18)
15. Er streckte unvermittelt die Arme aus und wollte der Mutter die Hände mit Gewalt vom Gesicht ziehen. (17)
16. Ganz besonders liebe ich stille Abende zu zweit, mit Wein, Musik, gutem Essen und dem Schein einer einzelnen Kerze. (19)
17. Wenn der Mann mit den Kindern nicht zurechtkommt, hat er ihre Liebe meiner Meinung nach auch nicht verdient. (18)
18. Am nächsten Tag fragte mich mein Vater, ob ich mitkomme, einen Strandkorb mit blauen Streifen zu besorgen. (17)
19. Langsam öffnete ich meine Augen, richtete mich im Bett auf, schlug die Decke weg und sah zur Uhr. (18)
20. Sie sah durch Bäume und Gestrüpp eine Wiese, auf der gelbe Tulpen und schöner weißer Flieder blühten. (17)
21. Die Sportler stellten rund achtzig Prozent der Weltmeister und errangen zwei Drittel aller möglichen Medaillen, zumeist in Gold oder Silber. (20)
22. Das Zimmer war von mir gewischt worden, jetzt mußte ich noch schnell die Fenster und den Spiegel putzen. (18)
23. Lange Zeit suchte ich nach dem Buch, bis ich es endlich in der Bibliothek jenseits der Stadtgrenze fand. (18)
24. Er sah sich jetzt davongetragen, mit großer Macht in die Nacht hinausgeschleudert wie ein Bündel Wäsche. (16)

25. Die Spur des Mädchens führte zu einem sehr großen Felsblock und verlor sich an dieser Stelle im Nichts. (18)
26. Das offene Fenster zitterte im Wind und dadurch, dass es plötzlich zuschlug, wurde ich wach und schrie vor Schreck auf. (20)
27. Die ruhige und harmonische Stimme meiner Mutter schien aus einer anderen, längst vergessenen Welt zu kommen. (16)
28. Im direkten Vergleich mit den westeuropäischen Ländern bot sich dem Betrachter 1941 in Ungarn ein trauriges Bild. (17)
29. Zur Hochzeit schenkte mir meine Mutter einen Schnellkochtopf, vor der Benutzung habe ich ein wenig Angst. (16)
30. Obwohl es erst Mittag war, war der Himmel dunkel, und wir mussten zum Lesen das Licht einschalten. (17)
31. Da kam der Junge in den Raum geschlendert, grüßte uns beiläufig und ließ sich auf einen Stuhl fallen. (18)
32. Es ist ratsam, die Kakteen und den Gummibaum mindestens einmal im Monat mit entkalktem Wasser zu besprühen. (17)
33. Nur dieser eine Tag blieb ihm in der fremden Stadt, um dort Denkmal und Kirche zu sehen. (17)
34. Mondschein und Sterne glänzten während des Gesprächs in den Fenstern jenseits der Straße und lenkten sie ab. (17)
35. Dann fügte ich vor dem Ende der Geschichte eine kleine Kunstpause ein, um sie auf die Folter zu spannen. (19)
36. Der Chef schreibt dann die Zahlen der Filialen auf und sagt entweder, das ist gut oder das ist schlecht. (19)
37. Heute musst du hier bleiben und unser Gast sein, es ist ja schon fast Abend und es schneit. (18)
38. Meine Freundinnen und ich kletterten lachend auf die hohe Bretterbühne und drehten uns bis zum Umfallen. (16)
39. Kurz vor der Stadt verlor er völlig die Gewalt über das Moped und fuhr in den Graben. (17)
40. Einen Augenblick vergaß Thomas vollkommen, dass der Vater noch am Fußende des Bettes stand und ihm zuhörte. (17)

A2. C-Test Deutsch Studien 2, 3 und 5

Ergänze bitte die Lücken in den folgenden Texten.

Beispiel: Die Bäume verlieren im Herbst ihre Blätter.

Text 1

Nach dem Kindergarten gehen alle Kinder im Alter von sechs oder sieben Jahren in die Grundschule. Dort bleiben sie vier Jahre. Danach gehen ca. 25 Prozent auf die Hauptschule, ca. 45 Prozent auf die Realschule und ca. 30 Prozent auf das Gymnasium. In vielen Bundesländern gibt es auch Gesamtschulen. Nach der Hauptschule kann man arbeiten und Geld verdienen oder eine Lehre machen. Nach dem Abi am Gymnasium kann man einen Beruf lernen oder an der Universität studieren.

Text 2

In sechs Stunden kann man per Straßenbahn quer durch das Ruhrgebiet fahren. 5,6 Millionen Menschen wohnen und arbeiten hier. Die Städte sind durch ein kompliziertes Netz von Straßenbahnen und Bus-Linien verbunden. Die längste Straßenbahn-Linie der Welt ist 115 km lang. Man kann sie von Düsseldorf bis nach Dortmund befahren. Kommen Sie mit, stehen Sie einkaufs; vergessen Sie das Umsteigen an der richtigen Haltestelle nicht! Unsere Fahrt beginnt in Düsseldorf.

Text 3

Zeitungsvolltexte via Computer sollen Blinden den Zugang zur täglichen Nachrichtenwelt verschaffen. Auf Initiative des Deutschen Blindenverbandes und der Stiftung Blindenanstalt wird ab dem Samstag in München im Rahmen der „Reha“ die Demonstrationsversion einer „Elektronischen Tageszeitung“ für Blinden gezeigt. Der Inhalt einer überregionalen Tageszeitung wird auf einem Fernsehkanal übertragen und kann von ausgesuchten Testpersonen über ein blindengerechtes Ausgabegerät am Computer gelesen werden.

Text 4

Der Streik der Transportunternehmen in Italien, der eine fast einwöchige Versorgungskrise vor allem in der Produktion, bei Nahrungsmitteln und Benzin auslöste, ist am Sonntag vorerst beendet worden. Regierung und Vertreter der rund 210.000 Transportunternehmen werden in den kommenden Wochen über die umstrittene Struktur des Güterverkehrs und Ausgleichszahlungen wegen erhöhter Kraftstoffpreise verhandeln. Bei einem Scheitern wollen die Transportunternehmen erneut streiken. Italiens Versorgung ist weitgehend von LKWs abhängig, die rund 80 Prozent aller Waren im Land transportieren. Die Regierung will einen großen Teil der Warenströme mittels einer Steuerreform auf die Schiene verlagern.

A3. Adaptierte Sätze aus Lesespannentest, Französisch

Die Wortzahl der Sätze wurde angepasst an die Auswahl der deutschen Sätze, die in Studie 3 verwendet wurden.

TEST 1 Il était un peu moins de midi trente quand son amie sortit lentement de l'abri.

TEST 2 Elle attendit que la brume se lève et put alors admirer le lac et la vallée.

TEST 3 Ses nerfs étaient tendus car on lui avait répété que son adversaire était en forme.

TEST 4 Le prêtre traversa la cour aux murs blanchis et referma la lourde porte de la tour.

1. Le navire se mit à prendre l'eau car sa coque avait été percée la veille par un rocher.
(18)

2. Dès qu'elle fut arrivée sur la rive droite du fleuve, la petite troupe se disposa en cercle.
(17)

3. Le flacon argenté contenait un liquide bleu clair et frais, au goût de miel et de banane confite. (18)

4. A ce moment, le train en provenance de Varsovie entra en gare et mon ami abandonna sa lecture avec regret. (20)

5. Le lendemain matin, il retrouva la clef de sa maison parmi les verres, les cotillons et les bouteilles. (18)

6. Depuis deux longs jours la petite baraque en planches et au toit vitré nous protégeait de la tempête de sable. (20)

7. Il se fit passer pour un inspecteur en civil et entra donc avec beaucoup de facilité dans la banque. (19)

8. A trois heures précises, le troupeau de chèvres s'engouffra sous le portail de la grande ferme, accompagné de son berger. (19)

9. Le barbier s'approcha de mon visage tout en aiguisant un vieux rasoir sur le cuir de son pantalon usé. (19)

10. Il a passé toutes ses vacances d'été à chercher les trésors enfouis au plus profond de l'océan. (17)

11. La jeune adolescente, énervée, s'appuya contre la voiture et la frappa trois fois des poings avec violence. (17)

12. Malgré tout son courage et sa persévérance, la servante ne parvenait pas à accomplir son dur labeur. (17)

13. Rappelé en urgence par un télégramme obscur, le baron fit le plus pénible des cauchemars la nuit suivante. (18)

14. Le conducteur de la moto portait un casque et une combinaison de cuir mais jamais de lunettes. (17)
15. Il fut ainsi conduit hors de notre petit pays, au milieu d'une foule de bourgeois en colère. (17)
16. Le mercredi, sa mère glissait toujours dans sa poche une mince galette de blé, pour manger à l'école. (18)
17. Le tonnerre grondait sauvagement sur le chantier du centre de loisirs et la foudre tomba sur le toit. (18)
18. Elle avait l'impression d'être constamment épiée lorsqu'elle marchait dans la rue et éprouvait un profond malaise. (17)
19. Lors du contrôle technique, le nouveau mécanicien lui avait conseillé de vérifier plus souvent le niveau d'huile. (17)
20. Elle rassembla ses maigres affaires et partit sans se retourner loin de ce village pour aller vivre à Paris. (19)

A4. C-Test Französisch

Ergänze bitte die Lücken in den folgenden Texten.

Beispiel: Die Bäume verlieren im Herbst ihre Blätter.

Text 1

Le paysan tient à fois du propriétaire et de l'ouvrier. Le cara _____ manuel d' son tra _____, l'eff _____ physique l'appar _____ à l'ouv _____, mais l' _____ propriété l'e _____ distingue radica _____. Rien d' _____ plus évident dans n' _____ villages d' _____ Nord, e' _____ bordure d' _____ pays min _____. Aux ye _____ du pay _____, le min _____ est u' _____ roi.

Text 2

La première tendance de la fête, c'est de plus en plus celle du repli sur soi, sur les petits groupes, sur la famille. Les tradi _____, les ri _____, demeurent tr' _____ forts po' _____ célébrer ch' _____ des occa _____ essentielles d' _____ l'exis _____: le bap _____, le mar _____, l'enter _____, sans oub _____, survivance pure _____ sociale q' _____ n'a pl' _____ guère d' _____ racines relig _____, la première communion, c'est-à- _____ le pas _____ de l'enfance à la vie adulte.

Text 3

"L'université fabrique des chômeurs." Cette affir _____ en fo' _____ de slo _____, est, po' _____ beaucoup, e' _____ passe d' _____ devenir u' _____ évid _____. Il y aur _____ trop d'étud _____, trop d' _____ diplômés, trop d'intell _____. De nomb _____ exemples, pl' _____ ou mo' _____ spectaculaires, vien _____ à l'ap _____ de ce _____ thèse. E' _____ pourtant il faut y regarder de plus près.

Text 4

Le libre choix de son médecin est pour le malade un droit; l'exer _____ de c' _____ droit do' _____ être faci _____. La vol _____ du mal _____ doit touj _____ être resp _____ dans to' _____ la mes _____ du poss _____. Lorsque l' _____ malade e' _____ hors d'ét _____ d'expr _____ son consen _____ aux so' _____ qu'o' _____ se pro _____ de l' _____ donner, ses proches doivent, sauf urgence ou impossibilité, être prévenus ou informés.

Text 5

La peur: elle baigne l'atmosphère banlieusarde, dramatise son ennui. C'e' _____ un phén _____ qui rel _____ plutôt d' _____ la psyc _____ collective q' _____ de l' _____ peur prov _____ par d' _____ faits ré _____. Il y a cer _____ des cass _____ de voi _____, des agres _____ commises sou _____ par des toxic _____, des vo' _____, des cambri _____, des бага _____, mais ri _____ de dramatique.

A5. Verständnisfragen Bedingung "Verstehen" Studie 4

1. Verwende ich das Geschenk meiner Mama häufig? n
2. Waren wir fröhlich? j
3. Wurde ich angeschrien? n
4. War es draußen dunkel? j
5. Will mein Papa mit mir einkaufen gehen? j
6. Hatte er viel Zeit? n
7. War sie aufmerksam an der Unterhaltung beteiligt? n
8. Sah seine Mutter ihm in die Augen? n
9. Brauchten wir eine Lampe? j
10. War der Vater verärgert? j
11. War Thomas die Anwesenheit seines Vaters immer bewusst? n
12. Gibt es Schwierigkeiten in der Schule? j
13. Kam er wohlbehalten in der Stadt an? n
14. War es Frühling? j
15. Gab es 1941 in Ungarn viel Schönes zu sehen? n
16. Ist es Winter? j
17. Suchte man das Mädchen im Wasser? n
18. Hatte er Spaß bei seiner Arbeit? n
19. Wurde Elke nass? j
20. Ist der Mann selbst schuld? j
21. Kaufte ich das Buch im Internet? n
22. Blieb er den ganzen Morgen draußen? j
23. Können in der Stadt 1000 Patienten stationär behandelt werden? j
24. Gab der Junge uns die Hand? n
25. Hatte es geregnet? j
26. Machte Rainer einen Ausflug mit seiner Freundin? n
27. Hat Klaus Angst vor dem Direktor? n
28. Konnte Anna im Urlaub nicht abschalten? j
29. Ist der Boden schon sauber? j
30. Schaute ich aus dem Fenster? n
31. Erzählte ich schnell bis zum Ende? n
32. Gibt der Chef am Ende ein Urteil ab? j
33. Versuchte man, ihn ins Freie zu bringen? j
34. Verschonte er die Schüler mit seinen Prinzipien? n
35. Gehe ich am Liebsten auf Parties? n
36. Sollte das Wasser für die Kakteen entkalkt sein? j
37. Waren die Sportler im Wettkampf erfolgreich? j
38. Flog das Fenster auf? n
39. Wohne ich in der Nähe des Waldes? n
40. Gab es ein Unwetter? j

A6. Sätze für Studie 5

5. Und auch seine Prinzipien übertrug er konsequent auf die Schüler und die ganze Schule, die er seit August leitete.
8. Wenn irgend möglich, setze ich mich in die Straßenbahn, fahre ich nach außerhalb und gehe ein bisschen im Wald spazieren.
10. Die Stadt verfügt seit mehreren Jahren über ein modernes Krankenhaus mit rund 1000 Betten und bester technischer Ausrüstung.
11. Da das Konzert im Freien stattfand, musste es wegen des einsetzenden Sturms abgebrochen und die Zuschauer in Sicherheit gebracht werden.
13. Er fertigte sein letztes Werkstück kurz vor Feierabend ohne jegliches Interesse an und gab es gleichgültig zur Kontrolle.
14. In pechschwarzer Nacht stand Elke in ihren ältesten Mantel gehüllt an der Kreuzung unter einem Baum im Regen.
16. Ganz besonders liebe ich stille Abende zu zweit, mit Wein, Musik, gutem Essen und dem Schein einer einzelnen Kerze.
18. Am nächsten Tag fragte mich mein Vater, ob ich mitkomme, einen Strandkorb mit blauen Streifen zu besorgen.
19. Langsam öffnete ich meine Augen, richtete mich im Bett auf, schlug die Decke weg und sah zur Uhr.
20. Sie sah durch Bäume und Gestrüpp eine Wiese auf der gelbe Tulpen und schöner weißer Flieder blühten.
21. Die Sportler stellten rund 80 Prozent der Weltmeister und errangen zwei Drittel aller möglichen Medaillen, zumeist in Gold oder Silber.
22. Das Zimmer war von mir gewischt worden, jetzt musste ich noch schnell die Fenster und den Spiegel putzen.
23. Lange Zeit suchte ich nach dem Buch, bis ich es endlich in der Bibliothek jenseits der Stadtgrenze fand.
25. Die Spur des Mädchens führte zu einem sehr großen Felsblock und verlor sich an dieser Stelle im Nichts.
26. Das offene Fenster zitterte im Wind und dadurch, dass es plötzlich zuschlug, wurde ich wach und schrie vor Schreck auf.
30. Obwohl es erst Mittag war, war der Himmel dunkel, und wir mussten zum Lesen das Licht einschalten.

31. Da kam der Junge in den Raum geschlendert, grüßte uns beiläufig und ließ sich auf einen Stuhl fallen.

36. Der Chef schreibt dann die Zahlen der Filialen auf und sagt entweder, das ist gut oder das ist schlecht.

37. Heute musst Du hier bleiben und unser Gast sein, es ist ja schon fast Abend und es schneit.

40. Einen Augenblick vergaß Thomas vollkommen, dass der Vater noch am Fußende des Bettes stand und ihm zuhörte.

B Fragebögen zur Spracherwerbsgeschichte

B1. Fragebogen Kinder Studie 1

Sprachgeschichte von _____ (Name des Kindes)

Um die Ergebnisse der Studie richtig einordnen zu können, brauche ich ein paar Informationen über Ihr Kind und die Sprachen, die Ihr Kind spricht oder versteht. Bitte versuchen Sie, die folgenden Fragen möglichst genau zu beantworten:

1) Wie alt ist Ihr Kind? ____ Jahre und ____ Monate

2) Die erste Sprache, die Ihr Kind gesprochen und verstanden hat ist:

☐ Deutsch ☐ _____

3) Sind bei Ihrem Kind Sprachstörungen bzw. Entwicklungsstörungen in dieser Sprache bekannt?

Wenn Sie oben **nicht** "Deutsch" angekreuzt haben, beantworten Sie bitte auch die folgenden Fragen:

4) Welche Sprache sprechen Sie zuhause überwiegend mit Ihrem Kind?

Mutter: _____

Vater: _____

4a) Hat Ihr Kind **ältere** Geschwister? Welche Sprache(n) sprechen die Geschwister überwiegend untereinander?

5) Seit wie vielen Monaten besucht Ihr Kind eine deutschsprachige Kindertagesstätte? ____

6) Wie viele Stunden pro Tag verbringt Ihr Kind durchschnittlich in der Kita? ____

7) Wie viele Tage pro Woche besucht Ihr Kind durchschnittlich die Kita? ____

8) Spricht Ihr Kind außerhalb der Kindertagesstätte Deutsch? In welchem Umfang (bitte schätzen Sie die Stunden pro Woche)?

9) Kam Ihr Kind vor dem Besuch der deutschsprachigen Kindertagesstätte mit der deutschen Sprache in Kontakt? Wie, ab welchem Alter und in welchem Umfang?

Vielen Dank!!

B2. Fragebogen Studenten (Studien 2 und 5)

VP#

Um die Daten des Experimentes statistisch auswerten zu können, brauchen wir noch ein paar Informationen von Dir. Deine Antworten auf die folgenden Fragen werden natürlich anonym behandelt.

1. Deutsch ist meine

___ Muttersprache ___ Zweitsprache ___ Dritt- oder Viertsprache

1.a Wenn **Deutsch Deine Muttersprache ist**, sprichst Du noch andere Sprachen?

___nein ___ja, und zwar:

A _____ seit meinem _____ Lebensjahr Wo? ___zu Hause ___Schule/Uni/Beruf

B _____ seit meinem _____ Lebensjahr Wo? ___zu Hause ___Schule/Uni/Beruf

1.b Wenn **Deutsch nicht Deine Muttersprache ist**:

Was ist Deine Muttersprache? _____

Sprichst Du außerdem noch Sprachen? Welche? _____

Wann hast Du angefangen, Deutsch zu lernen? mit _____ Jahren

Hast Du Deutsch hauptsächlich im Alltag oder im Unterricht gelernt? (Beispiel: „zu 30 % in der Schule und zu 70 % im privaten Umfeld“)

Welche Sprache sprichst/sprachst Du

mit Deiner Familie? _____

mit Deinen Freunden? _____

in der Schule? _____

In welcher Sprache hast Du erstmals Schreiben gelernt? _____

Wie lange lebst Du insgesamt in Deutschland bzw. einem deutschsprachigen Land?

Wie beurteilst Du Deine Kenntnisse in Deutsch: (bitte ankreuzen)

Grundkenntnisse

Muttersprachler

1 2 3 4 5 6

2. Wie alt bist Du jetzt? _____

3. Bist Du weiblich _____ oder männlich _____?

Vielen Dank für Deine Teilnahme!!

B3. Fragebogen Schüler Studie 3

Persönlicher Code:

Geburtsdatum der Mutter (TTMM), erste zwei Buchstaben des Vornamens der Mutter:

--	--	--	--	--	--

Fragebogen zur Sprachgeschichte

1) Welche Sprache ist Deine Muttersprache?

- ☐ Deutsch
- ☐ Französisch
- ☐ Deutsch und Französisch bilingual (beide Sprachen seit Geburt)
- ☐ Andere: _____

3) Ab welchem Alter hattest Du Kontakt zu Deiner Partnersprache? _____

2) In welchem Alter hast Du begonnen, Deine Partnersprache (intensiv) zu lernen? _____

4) Welche Sprachen sprechen Deine Eltern? Welche Sprache sprichst Du überwiegend mit Deinen Eltern?

	Mutter	Vater
Muttersprache		
ggf. Zweitsprache		
Was spricht Ihr miteinander?		

5) Hast Du ältere Geschwister? Wenn ja, welche Sprache spricht ihr überwiegend miteinander?

6) Wie schätzt Du Deine Sprachfertigkeit in Deinen beiden Sprachen ein?

Deutsch: *Grundkenntnisse* 1-----2-----3-----4-----5-----6 *Muttersprache*
Französisch: 1-----2-----3-----4-----5-----6

7) Welche Sprache sprichst Du überwiegend?

	<i>Französisch</i>	<i>Deutsch</i>
in der Schule:	ca. _____ % der Zeit	ca. _____ % der Zeit
in Deiner Familie	ca. _____ % der Zeit	ca. _____ % der Zeit
mit Freunden	ca. _____ % der Zeit	ca. _____ % der Zeit

Vielen Dank!! Fragebogen bitte am Tag der Studie mitbringen oder mit der Einverständniserklärung zur Teilnahme abgeben.

C Instruktionen Studien 4 und 5

Instruktionen Studie 4

Instruktion wörtliches Behalten

Liebe/r Teilnehmer/in,

im folgenden Experiment wirst Du Sätze lesen und sollst Sie dann wiedergeben.

Die Sätze werden Dir Wort für Wort am Bildschirm präsentiert. Um das nächste Wort zu sehen, musst Du die Leertaste drücken. Jedes Wort wird Dir nur einmal angezeigt, Du kannst also nicht im Satz zurückspringen. Lies so genau und schnell, wie es Dir möglich ist.

Nach jedem Satz hast Du Zeit, ihn auf dem Ergebnisblatt aufzuschreiben.

Schreibe den Satz wörtlich und so genau auf, wie Du Dich an ihn erinnerst. Schreibe alles auf, was Du erinnerst. Achte dabei darauf, dass die Sätze, die Du aufschreibst grammatisch korrekt sind (ergänze zur Not mit eigenen Wörtern).

Danach drücke wieder die Leertaste, um den nächsten Satz zu starten.

Zunächst werden Dir drei Sätze zum Üben präsentiert. Wenn Du Fragen hast, sprich bitte die Versuchsleiter an.

Instruktion inhaltliches Behalten

Liebe/r Teilnehmer/in,

im folgenden Experiment wirst Du Sätze lesen und sollst Sie dann wiedergeben.

Die Sätze werden Dir Wort für Wort am Bildschirm präsentiert. Um das nächste Wort zu sehen, musst Du die Leertaste drücken. Jedes Wort wird Dir nur einmal angezeigt, Du kannst also nicht im Satz zurückspringen. Lies so genau und schnell, wie es Dir möglich ist.

Nach jedem Satz hast Du Zeit, ihn auf dem Ergebnisblatt aufzuschreiben.

Dabei ist nur der Inhalt des Satzes wichtig, es ist egal, ob Du ihn in den gleichen Worten oder mit Deinen eigenen Worten wiedergibst. Versuche, den Inhalt so vollständig es Dir möglich ist aufzuschreiben. Achte darauf, dass die Sätze, die Du aufschreibst, grammatisch korrekt sind.

Danach drücke wieder die Leertaste, um den nächsten Satz zu starten.

Zunächst werden Dir drei Sätze zum Üben präsentiert. Wenn Du Fragen hast, sprich bitte die Versuchsleiter an.

Instruktion Verstehen

Liebe/r Teilnehmer/in,

im folgenden Experiment wirst Du Sätze lesen.

Die Sätze werden Dir Wort für Wort am Bildschirm präsentiert. Um das nächste Wort zu sehen, musst Du die Leertaste drücken. Jedes Wort wird Dir nur einmal angezeigt, Du kannst also nicht im Satz zurückspringen. Lies so genau und schnell, wie es Dir möglich ist.

Nach jedem Satz erscheint eine Frage, die sich auf den Satz bezieht.
Beantworte die Frage mit "ja" oder "nein", indem Du die entsprechend markierte Taste drückst. Beantworte die Fragen jeweils so schnell und korrekt wie möglich.
Danach drücke wieder die Leertaste, um den nächsten Satz zu starten.

Zunächst werden Dir drei Sätze zum Üben präsentiert. Wenn Du Fragen hast, sprich bitte die Versuchsleiter an.

Instruktionen Studie 5

Instruktion wörtliches Behalten

Liebe/r Studienteilnehmer/in,
im folgenden Experiment wirst Du Sätze anhören, setze dazu bitte die Kopfhörer auf.
Deine Aufgabe ist es, die Sätze wiederzugeben. Fange damit nicht eher an, als der Satz zu Ende ist.
Nach jedem Satz hast Du Zeit, ihn auf dem Ergebnisblatt aufzuschreiben.
Schreibe den Satz wörtlich und so genau auf, wie Du Dich an ihn erinnerst. Schreib alles auf, was Du erinnerst. Achte dabei darauf, dass die Sätze, die Du aufschreibst grammatisch korrekt sind (ergänze zur Not mit eigenen Wörtern).
Danach drücke die Leertaste, um den nächsten Satz zu starten.
Um Dich mit dem Ablauf vertraut zu machen, werden Dir zwei Probesätze dargeboten. Wenn Du nun keine Fragen mehr hast, kannst Du damit beginnen, indem Du die Leertaste drückst.

Instruktion inhaltliches Behalten

Liebe/r Studienteilnehmer/in,
im folgenden Experiment wirst Du Sätze anhören, setze dazu bitte die Kopfhörer auf.
Deine Aufgabe ist es, die Sätze wiederzugeben. Fange damit nicht eher an, als der Satz zu Ende ist.
Nach jedem Satz hast Du Zeit, ihn auf dem Ergebnisblatt aufzuschreiben.
Dabei ist nur der Inhalt des Satzes wichtig, es ist egal, ob Du ihn in den gleichen Worten oder in Deinen eigenen Worten wiedergibst. Versuche, den Inhalt so vollständig es Dir möglich ist, aufzuschreiben. Achte darauf, dass die Sätze, die Du aufschreibst, grammatisch korrekt sind.
Danach drücke die Leertaste, um den nächsten Satz zu starten.
Um Dich mit dem Ablauf vertraut zu machen, werden Dir zwei Probesätze dargeboten. Wenn Du nun keine Fragen mehr hast, kannst Du damit beginnen, indem Du die Leertaste drückst.

Im Folgenden sind Kopien von Quelldokumenten angehängt, die zitiert wurden, aber nicht in üblicher Form (z.B. in Bibliotheken) zur Verfügung stehen.

E1.



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Georg-Elias-Müller-Institut für Psychologie
Abteilung Pädagogische Psychologie
und Entwicklungspsychologie
Prof. Dr. Marcus Hasselhorn
Waldweg 26
37073 Göttingen

An das
Ministerium für Kultus, Jugend und Sport
Baden-Württemberg
z.H. Frau Christa Engemann
Postfach 10 34 42
70029 Stuttgart

Sekretariat
Tel. +49(0) 551 / 39-9289
Fax +49 (0) 551 / 39-9322
rzsoerne@uni-goettingen.de

Göttingen, 13.08.2007

**Betr.: Expertise über derzeit verfügbare wissenschaftlich fundierte
Verfahren einer umfassenden Sprachstandsdiagnostik bei Vier- und
Fünfjährigen**

Sehr geehrte Frau Engemann,

gerne komme ich hiermit Ihrer Bitte nach, eine gutachterliche Stellungnahme über derzeit im deutschsprachigen Raum verfügbare Verfahren zur umfassenden Sprachstandsdiagnostik bei Kindern im Alter von 4 und 5 Jahren abzugeben. Die folgenden Ausführungen stehen im Zusammenhang mit der Diskussion um flächendeckende verbindliche Sprachstandsdiagnostik in Baden-Württemberg etwa 1½ Jahre vor der Einschulung, um darauf aufbauend gezielte Fördermaßnahmen durchzuführen. Eine allgemeine Expertise zu Sprachstandserhebungen für Kindergartenkinder und Schulanfänger wurde 2004 im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend am Deutschen Jugendinstituts in München von Fried (2004) erstellt. In dieser Expertise wird deutlich herausgestellt, dass die aktuell im deutschsprachigen Raum verfügbaren Verfahren sehr unterschiedliche Funktionen erfüllen.

Die meisten der derzeit in den bildungspolitischen Diskussionen gehandelten Verfahren sind reine Screening-Verfahren. Screening-Verfahren dienen der raschen und ökonomischen Identifikation von „Risikokindern“, d.h. von Kindern, bei denen mit mehr oder minder hoher Sicherheit der begründete Verdacht besteht, dass länger anhaltende gravierende Sprachdefizite bestehen. Zu den bekanntesten aktuellen Verfahren dieser Art gehören z.B. das in Baden-Württemberg entwickelte und erprobte Verfahren HASE. Solche Verfahren zur Sprachstandsdiagnostik bei Kindern sind vor allem gesundheitspolitisch interessant, wenn sie standardisiert und normiert sind und ihre prognostische Validität empirisch nachgewiesen ist (was leider nur in Ausnahmefällen – z.B. bei HASE – der Fall ist). Sie haben aber den Nachteil, dass sie keine Information darüber liefern, worin die Sprachentwicklungsrückstände im einzelnen bestehen und welche Fördermaßnahmen daher pädagogisch für die einzelnen Kinder angezeigt sind.

Will man eine möglichst umfassende und förderdiagnostisch einsetzbare Sprachstandsdiagnostik realisieren, so stehen prinzipiell zwei diagnostische Methoden zur Verfügung, die Analyse von Spontansprache sowie die Verwendung standardisierter Testverfahren. Die von Sprachwissenschaftlern bisweilen bevorzugte Methode der Spontansprachanalyse hat diagnostisch zwei große Nachteile. Zum einen erlaubt diese Methode keine Feststellung des gerade für die schulische Entwicklungsprognose so entscheidenden Sprachverstehens, zum anderen ist die Standardisierung bei der Erfassung von Spontansprache nur bedingt möglich, so dass die Ergebnisse unterschiedlicher Kinder nur begrenzt vergleichbar sind und so ein flächendeckender Einsatz mit dem Ziel, auf den Ergebnissen konkrete Fördermaßnahmen aufzubauen nicht wirklich gut erreicht wird.

Roos und Schöler (2007) sehen daher keine seriöse Alternative zu den testtheoretisch fundierten und bewährten Sprachentwicklungstests. Ähnlich ist hierzu auch die Einschätzung in der Expertise von Fried (2004) – eine umfassende und flächendeckend handhabbare Sprachstandsdiagnose sollte auf theoretisch begründeten und empirisch bewährten Testverfahren basieren, die sich hinsichtlich der wissenschaftlich verbindlichen Gütekriterien als alternativen vorliegenden Verfahren überlegen erwiesen haben.

Allerdings wird aus der Expertise von Fried (2004) auch deutlich, dass es derzeit kein Verfahren gibt, das in jeder Hinsicht als exzellent geeignet eingestuft werden könnte. Aus diesem Grunde ist die vorliegende Expertise darum bemüht, das gegenwärtig relativ geeignetste Verfahren zu identifizieren. Dazu habe ich unter den verfügbaren diagnostischen Verfahren zur eher umfassenden Erfassung des Sprachstandes bei Vier- und Fünfjährigen geprüft, wie gut sie hinsichtlich der folgenden Hauptkriterien abschneiden: (1) Theoretische Fundierung, (2) Objektivität (Standardisierung), (3) Reliabilität, (4) Validität, (5) Normierung und (6) Ökonomie.

ad (1) Theoretische Fundierung

In Anlehnung an Weinert (2006; vergl. auch Roos & Schöler, 2007) sollte eine Diagnostik mit dem Anspruch, die Sprachkompetenz von Kindern einigermaßen umfassend zu erheben, folgende Sprachkomponenten berücksichtigen:

- *prosodisch-rhythmische Kompetenzen* (Rhythmische Gliederung, Betonungen, Dehnungen, Intonationskontur)
- *phonologisch-phonetische Kompetenzen* (Erkennen und Nutzung bedeutungsdifferenzierender Lautkategorien und deren phonotaktischen Kombinationsregeln)
- *morphologische Kompetenzen* (Fähigkeit zur Wortbildung aus bedeutungstragenden Einheiten)
- *syntaktische Kompetenzen* (Beherrschen formaler Kategorien von Wortordnungen)
- *lexikalisch-semantische Kompetenzen* (Verstehen der Bedeutungsstruktur von Wörtern und Sätzen)
- *pragmatische Kompetenzen* (Nutzung der Regeln der Sprachverwendung einschließlich Sprechakten)

Für den Alterbereich von 4;6 bis 5;11 Jahren erfüllen der HSET (Heidelberger Sprachentwicklungstest, Grimm & Schöler, 1991), der KISTE (Kindersprachtest für das Vorschulalter, Häuser, Kaselke & Scheidereiter, 1994), der SETK 3-5 (Sprachentwicklungstest für Kinder, Grimm, Aktas & Frevert, 2001) sowie – mit

Einschränkung – der MSVK (Marburger Sprachverständnistest für Kinder, Elben & Lohaus, 2000) die Anforderung, diese Teilkompetenzen weitgehend abzudecken.

ad (2) Objektivität (Standardisierung)

Ein diagnostisches Verfahren gilt dann als objektiv, wenn das diagnostische Urteil über die zu beurteilende Person nicht von der Person des Beurteiler beeinflusst wird. Man unterscheidet insgesamt drei Quellen möglicher Objektivitätsmängel: die Durchführung, die Auswertung und die Interpretation. Durch Standardisierung der Verfahren und Schulung der Testanwender werden die hiermit verknüpften Mängel minimiert. Hinsichtlich dieses Merkmales gibt es keine nennenswerte Unterschiede zwischen den vier nach dem ersten Auswahlkriterium überhaupt noch in Frage kommenden Verfahren.

ad (3) Reliabilität

Das Güte Merkmal der Reliabilität bezieht sich auf die Zuverlässigkeit bzw. Genauigkeit des diagnostischen Urteils. Von einem reliablen diagnostischen Verfahren erwartet man, dass es präzise und exakt misst. Empirisch lässt sich dies über sogenannte Reliabilitätskoeffizienten ermitteln, die theoretisch Werte zwischen 0 und 1 annehmen können, wobei ein Wert von 1 dem hypothetischen Idealfall der vollkommen von Messfehlern freien Messung entspricht.

Die berichteten empirischen Reliabilitäten für drei der vier hier in der engeren Wahl befindlichen Verfahren sind zufriedenstellen bis gut. Je nach Subtest variieren sie zwischen .74 und .95 für den HSET, zwischen .72 und .89 für den SETK 3-5 und zwischen .72 und .90 für die Kiste. Lediglich beim MSVK fallen Einzelteile unter ein akzeptierbares Reliabilitätsniveau (.35 bis .88).

ad (4) Validität

Das Güte Merkmal der Validität bezieht sich auf die Gültigkeit des diagnostischen Verfahrens, also darauf, wie gut das Verfahren das erfasst, was es zu messen beansprucht. Ähnlich wie bei der Reliabilität handelt es sich bei der Validität um ein Güte Merkmal, das nur graduell zu bestimmen ist. Bei der empirischen Bestimmung der Validität von Verfahren zur Sprachstandsdiagnose bei jüngeren Kindern ist vor allem die prognostische Validität von Bedeutung. Bei ihrer Bestimmung geht es um die Frage, wie gut aufgrund der diagnostischen Befunde durch das Verfahren relevante spätere Ereignisse (z.B. Sprachstörungen, Schriftsprachschwierigkeiten) vorhersagbar sind. Empirische Untersuchungen zur prognostischen Validität liegen bisher nur zum HSET und zum SETK 3-5 vor, wobei die zum SETK 3-5 noch unveröffentlicht (Dissertation der Mitautorin Frevert) sind. Sie sind allerdings nicht ausreichend, um die prognostische Validität abschließend zu bewerten.

ad (5) Normierung

Normwerte sind erforderlich, um individuelle Ergebnisse in einem diagnostischen Verfahren auf der Basis der in einer geeigneten Referenzpopulation erbrachten Leistungen zu bewerten. Für alle vier betrachteten Verfahren liegen empirisch bestimmte Normwerte vor, die als befriedigend eingestuft werden können (Normierungsstichproben zwischen 500 und 1000 Kindern).

ad (6) Ökonomie

Die Wirtschaftlichkeit gehört zu den Nebengütekriterien diagnostischer Verfahren, die insbesondere dann von Bedeutung sind, wenn sich mehrere ansonsten in ihrer Leistungsfähigkeit nicht groß unterscheidende Verfahren auf dem Markt befinden. Für ein flächendeckend zum Einsatz kommendes Verfahren ist vor allem der Zeitfaktor, der für die Testdurchführung zu veranschlagen ist, entscheidendes Merkmal der Ökonomie. Hinsichtlich dieses Merkmals ist der SETK 3-5 das beste der in Frage kommenden Verfahren (Durchführungszeit 20 bis 30 Minuten), gefolgt von MSVK (35 bis 40 Minuten), KISTE (35 bis 50 Minuten). Der HSET ist mit Abstand das zeitlich aufwendigste der in Frage kommenden Verfahren (Durchführungszeit ca. 70 Minuten).

Conclusio

Zusammenfassen lässt sich festhalten, dass für den Zweck eines flächendeckenden Einsatzes für alle die Kinder, die im Screening der Schuleingangsuntersuchung (in Baden-Württemberg durch HASE) als potentiell sprachbeeinträchtigt identifiziert wurden, der SETK 3-5 in der Summe das derzeit beste und damit zu empfehlende Verfahren darstellt. Nichtsdestotrotz sollte in den nächsten Jahren die weitere Entwicklung in der einschlägigen Literatur verfolgt werden, um gewahrt zu werden, ob der SETK 3-5 durch ein alternatives Verfahren im ersten Gesamtrangplatz abgelöst wird.

Literatur

- Elben, C.E. & Lohaus, A. (2000). *Marburger Sprachentwicklungstest für Kinder MSVK*. Göttingen: Hogrefe.
- Fried, L. (2004). *Expertise zu Sprachstandserhebungen für Kindergartenkinder und Schulanfänger: Eine kritische Betrachtung*. München: Deutsches Jugendinstitut.
- Grimm, H., Aktas, M. & Frevert, S. (2001). *Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)*. Göttingen: Hogrefe.
- Grimm, H. & Schöler, H. (1991). *Der Heidelberger Sprachentwicklungstest H-S-E-T* (2. Aufl.). Göttingen: Hogrefe.
- Häuser, D., Kasielke, E. & Scheidereiter, U. (1994). *Kindersprachtest für das Vorschulalter KISTE*. Weinheim: Beltz.
- Roos, J. & Schöler, H. (2007). Sprachentwicklungsdiagnostik mittels standardisierter Tests. In H. Schöler & A. Welling (Hrsg.), *Sonderpädagogik der Sprache* (S. 531-550). Göttingen: Hogrefe.
- Weinert, S. (2006). Sprachdiagnostik. In F. Petermann & M. Eid (Hrsg.), *Handbuch der Psychologischen Diagnostik* (S. 567-575). Göttingen: Hogrefe.



(Prof. Dr. Marcus Hasselhorn)

**Gewerkschaft Erziehung und Wissenschaft
Hauptvorstand**



26. April 2007

GEW: "Tests sofort stoppen"

Bildungsgewerkschaft zu Kita-Sprachprüfungen in Nordrhein-Westfalen

Frankfurt a.M. – Die Gewerkschaft Erziehung und Wissenschaft (GEW) hat gefordert, die Sprachtests für vierjährige Kinder in Nordrhein-Westfalen sofort zu stoppen. „Die schlechten Testergebnisse sind der Beleg, dass der Delfin-Test völlig ungeeignet ist. Sie sagen nichts darüber aus, wie gut Kinder die deutsche Sprache beherrschen.“, sagte Norbert Hocke, für Jugendhilfe und Sozialarbeit verantwortliches GEW-Vorstandsmitglied.

„Die Versuchsanordnung des Tests ist so angelegt, dass man nur herausfinden kann, ob Befehl und Gehorsam funktionieren“, stellte Hocke fest, „über Sprachfähigkeit wird man nichts erfahren.“ Kinder entwickelten Freude am Sprechen, wenn sie etwas von sich mitteilen wollen. Von einem vierjährigen Kind jedoch zu verlangen, dass es einen Satz wie „Der Apfel wird von Lena schmatzend gegessen“ korrekt nachspricht, zeuge lediglich davon, wie wenig sich die Prüfer in die Lebens- und Gedankenwelt kleiner Kinder hineinversetzt haben. Der Test verängstige die Mädchen und Jungen, statt ihnen zu helfen. Zudem bestehe die Gefahr, dass die sinnvollen Ansätze zur sprachlichen Integration von Kindern aus Migrantenfamilien mit dieser Art standardisierter Tests zunichte gemacht wird.

„Kindertagesstätten sind keine vorgezogenen Grundschulen. Sie arbeiten nach eigenen, kindgemäßen und altersgerechten Bildungsplänen. In der frühkindlichen Pädagogik ist längst bekannt, dass Kinder dann lernen, wenn der Inhalt Sinn und Bedeutung hat. Vorgegebenen Unsinn zu reproduzieren, löst bei den Kindern nur Verweigerung aus“, betonte Hocke. Das Testverfahren dürfe jetzt nicht in die zweite Runde gehen. „Die Förderung der Sprachkompetenz muss den hierfür ausgebildeten Erzieherinnen in den Kindertagesstätten übertragen werden“, sagte der GEW-Sprecher. „Um alle Kinder individuell und optimal fördern zu können, müssen die Rahmenbedingungen verbessert werden: In Gruppen mit 25 Kindern bei einem Stellenschlüssel von 1,5 ausgebildeten Erzieherinnen ist alles Engagement zum Scheitern verurteilt.“

Info: Im März 2007 mussten sich in Nordrhein-Westfalen alle Kinder, die im Alter von vier Jahren eine Kindertagesstätte besuchen, dem Sprachtest „Delfin 4“ unterziehen. Von den 145.000 untersuchten Mädchen und Jungen sind 95.000 durchgefallen. 62.000 Kinder müssen im Mai in eine zweite Testrunde. Der Sprachtest macht die Vorgabe, Kinder in Viergruppen in einem dreißigminütigen Verfahren nach strengen Regeln zu prüfen. Die Kinder durften nur vorgegebene Sätze und Wörter, auch solche ohne Sinn (z.B. „Der grüne Stuhl hüpfte fröhlich über die kalte Sonne.“) nachsprechen. Freies Sprechen war strikt verboten.



Delfin4 – ein Test geht baden

160.000 Kinder haben in diesen Wochen einen „Besuch im Zoo“ auf einem Spielbrett erlebt, mit dem ihre Sprachentwicklung getestet werden sollte. Bei über 40% der 3 ½ - 4 ½ -jährigen Kinder ergab der Test Auffälligkeiten, die in einer 2. Stufe nachgetestet werden sollen.

Nach vielen Erfahrungsberichten von mit dem Test beauftragten Grundschullehrerinnen und Erzieherinnen stellt die GEW fest:

1. Der Test ist völlig ungeeignet für die Zielgruppe.
 - Die Testdauer von 45-60 Minuten überfordert die Konzentrationsfähigkeit der Kinder.
 - Das Thema/ die Bilder des Spielplans entsprechen nicht der Erfahrung vieler Kinder, der Spielplan ist unübersichtlich und nicht motivierend.
 - Viele Kinder verweigern sich auf Grund der ungewohnten Testsituation, jede spontane Äußerung muss abgeblockt werden.
 - Die Testsprache schüchtert die Kinder ein, da sie nicht dem gewohnten Sprachgebrauch entspricht.
 - Unsinnwörter und -sätze werden eher von spracharmen Kindern nachgesprochen als von sprachbegabten („das gibt's doch gar nicht“).
2. Die Testdurchführung bindet über 3 Monate die Arbeitskraft von vielen Lehrerinnen und ErzieherInnen. Normaler Unterricht und vor allem Förderunterricht findet in dieser Zeit gerade in den Schulen nur noch eingeschränkt statt, in denen viele Kinder zu der Zielgruppe dieses Tests gehören.
3. Die bisher oft gute Zusammenarbeit Kita - Grundschule wird stark belastet, da die GS mit ihrer Beurteilung über die Höhe des „Kopfgeldes“ für die zu fördernden Kinder und damit über einen Teil der finanziellen Ausstattung der Kita entscheidet.
4. Der Testzeitpunkt ist problematisch, da viele der getesteten Kinder erst seit 6 Monaten in der Kita die deutsche Sprache erlernen.

Die GEW fordert die Landesregierung auf:

- Unterstützung und Ausbau der sprachentwicklungsbegleitenden Analyse der Kita und Streichung der umfangreichen Tests
- Kindergartenpflicht für alle Kinder mindestens 2 Jahre vor der Einschulung
- Fortbildung für die Erzieherinnen mit dem Schwerpunkt Sprachförderung
- Sprachförderung in allen Kitas nach einem integrativen Konzept statt Einrichtung von Sprachkursen für 4-jährige Kinder
- Qualifizierung der ErzieherInnen durch Fachhochschulausbildung
- Umsetzung der bewährten Konzepte wie Sismik und Havas zur Sprachförderung statt teure Entwicklung neuer Tests und Konzepte.

Die Sprachförderung in den Kitas sollte sich an folgenden Prinzipien orientieren:

- Die sprachlichen und kommunikativen Kompetenzen der Kinder werden gefördert, denn Sprache ist Teil der Identitätsbildung und Persönlichkeitsentwicklung
- Sprachförderung ist eine sozialpädagogische Querschnittsaufgabe für alle Erzieherinnen und erfordert sprachliche Sensibilität und Reflexivität in allen Teilen und bei allen Akteuren der Einrichtung und bezieht auch die Eltern mit ein.
- Die kindliche Neugier und Lernfreude und ihre Lust am Erforschen der Umwelt werden aufgegriffen und im Alltag für sprachliche Förderung genutzt.
- Sprachförderung erfolgt für alle Kinder.
- Die Chancen der Mehrsprachigkeit bei vielen Kindern werden besonders berücksichtigt.
- Sprachförderung braucht kleine Gruppen, damit alle Kinder zu Wort kommen können.
- Sprachförderung ist ein durchgehendes Konzept während der gesamten Kindergartenzeit, bei dem auch verschiedene Medien und Materialien, wie Bücher, Filme, PC genutzt werden.
- Die Erzieherinnen bereiten sich durch Fortbildungen auf diese Aufgabe vor und werden dabei von der Landesregierung unterstützt.

Für die GEW hat die ganzheitliche Förderung der Kinder im Elementarbereich, d.h. neben der sprachlichen auch die soziale und emotionale Entwicklung jedes Kindes eine hohe Priorität.

Kontakt zur Fachgruppe Grundschule, Stadtverband Köln:
Wolfgang Raabe, Tel.: 02203-51342

Gewerkschaft Erziehung und Wissenschaft · Stadtverband Köln
Hans-Böckler-Platz 1 · OGB-Haus · 50672 Köln · Telefon (0221) 516267 · Telefax (0221) 525446
www.gew-koeln.de · E-Mail: gew-koeln@netcologne.de

V.i.S.d.P.: Rixa Boms, Fachgruppe Grundschule, Gewerkschaft Erziehung und Wissenschaft, Landesverband NRW,
Nünningstr. 11, 45141 Essen



DEUTSCHE GESELLSCHAFT FÜR
SPRACHHEILPÄDAGOGIK E.V.



Deutscher Bundesverband
der akademischen
Sprachtherapeuten

Sprachstandserhebungen vor Schuleintritt

– Positionspapier Oktober 2007 –

Kirsten Kallmeyer

Bildungspolitische Meilensteine

Seit PISA 2000 werden die Forderungen immer konkreter.

Die Kultusministerkonferenz hat im Dezember 2001 als eines von sieben Handlungsfeldern zur Qualitätsentwicklung „Maßnahmen zur Verbesserung der Sprachkompetenz bereits im vorschulischen Bereich“ gefordert. Nach der Veröffentlichung von PISA 2003 haben Kultusministerkonferenz und Länder zudem vereinbart, die „Leistungen der Schülerinnen und Schüler mit schwachen Kompetenzen und der Jugendlichen mit Migrationshintergrund zu verbessern, also die Chancengerechtigkeit zu erhöhen“. Als eine Maßnahme in diesem zentralen Arbeitsbereich wurde im Februar 2005 festgelegt, dass „Sprachstandsdiagnosen und Sprachförderung im vorschulischen Bereich“ stattfinden sollen (vgl. BMBF 2006).

Reaktionen der Bundesländer

Die Umsetzung gestaltet sich ausgesprochen heterogen!

Entsprechend dieser bildungspolitischen Bestrebungen werden derzeit Sprachstandserhebungen vor Schuleintritt bereits in 4 deutschen Bundesländern (Berlin, Bremen, Niedersachsen, Nordrhein-Westfalen) verbindlich für alle Kinder umgesetzt und Bayern verfügt über ein verpflichtendes Programm für alle Kinder mit Migrationshintergrund. Weitere Bundesländer (z.B. Bayern, Brandenburg, Hessen) befinden sich bezüglich einer flächendeckenden Erhebung in der Planungsphase und/oder führen Sprachscreenings oder Sprachtests bei Vorschulkindern in Rahmen

umfassender Projekte durch (z.B. Baden-Württemberg, Hamburg, Hessen, Rheinland-Pfalz, Saarland, Sachsen, Schleswig-Holstein). In der Regel sollen so nicht nur Kinder mit Migrationshintergrund auf ihre sprachlichen Kompetenzen hin überprüft werden, auch deutschsprachig aufwachsende sollen ggf. eine adäquate vorschulische Sprachförderung erhalten.

dgs und dbs beziehen Position

Da die sprachspezifische Diagnostik, Förderung und Therapie von Vorschulkindern mit und ohne Migrationshintergrund zu den genuinen Kompetenzbereichen sprachspezifischer Wissenschaften, wie z.B. der Sprachheilpädagogik oder der Klinischen Linguistik zählen, ist es in Anbetracht der derzeit unzureichenden Umsetzung der bildungspolitischen Maßnahmen zur vorschulischen Sprachstandserhebung und Sprachförderung angezeigt, dass die *Deutsche Gesellschaft für Sprachheilpädagogik e. V.* (dgs) und der *Deutsche Bundesverband der akademischen Sprachtherapeuten* (dbs) Position beziehen und aus wissenschaftlicher Perspektive Vorschläge hinsichtlich einer sinnvolleren, effektiveren und ökonomischeren Umsetzung machen.

Kritik

Ressourcen werden verschwendet!

Wie derzeit die einzelnen Bundesländer ihren Auftrag erfüllen, Kinder mit sprachlichen Schwierigkeiten frühzeitig zu identifizieren und ggf. zu fördern, lässt sich auf verschiedenen Ebenen kritisch reflektieren. Die enorme Heterogenität der Vorgehensweisen macht hier eine umfassende Darstellung der optimierungswürdigen Aspekte unmöglich – folgende Punkte können aber als allgemeingültig angesehen werden:

Die Instrumente, anhand derer der Sprachstand der Kinder festgestellt werden soll, sind

- o *wenig aussagekräftig,*
- o *unzureichend evaluiert und*
- o *nicht ökonomisch.*

Die meisten durchführenden Pädagoginnen und Pädagogen (in der Regel handelt es sich um Angestellte in Kindertageseinrichtungen und in Regelschulen) sind *unzureichend professionalisiert*, um über eine Einschätzung zum Sprachstand hinaus konkrete Aussagen zum Sprachentwicklungsstand von Kindern zu machen bzw. Sprachtests durchzuführen.

Handlungsbedarf

Es muss sich etwas ändern!

Aus der Sicht der dgs und des dbs besteht also einerseits Handlungsbedarf bei der **Konzeption der Erhebungsverfahren** auf wissenschaftlichen Ebenen, wie der sorgfältigen Berücksichtigung linguistischer und diagnostischer Kriterien sowie der Mehrsprachigkeitsforschung, andererseits bei der **organisatorischen Umsetzung** der Maßnahmen, um wissenschaftliche wie bildungs- und gesundheitspolitische, finanzielle wie personelle Ressourcen besser zu nutzen und die Zuständigkeit von Fach- und Verantwortungsbereichen einzuhalten und besser zu vernetzen.

Handlungsstrategien

Wie können wir grundsätzlich sinnvoll intervenieren?

Grundsätzlich begrüßen *dgs* und *dfs* eine frühzeitige Intervention und halten eine *gestufte Vorgehensweise* bei der Feststellung des Sprachstandes für sinnvoll:

- (1) Durchführung einer verpflichtenden, gezielten **Langzeitbeobachtung** durch Erzieherinnen und Erzieher in Kindertageseinrichtungen anhand von Beobachtungs-instrumenten wie SSMIK und SELDAK, um Kinder mit offensichtlich regelhaftem Spracherwerb und ausreichenden Sprachkenntnissen zu identifizieren.
- (2) Alle Kinder, die keine Kindertageseinrichtung besuchen und diejenigen, bei denen durch die Langzeitbeobachtung nicht eindeutig ein regelhafter Spracherwerb festgestellt werden konnte, sollten an einer verpflichtenden **Sprachstandserhebung** teilnehmen. Eine ziel-führende Umsetzung dieser Feststellung des Sprachstandes erfordert zwingend eine Professionalisierung des Personenkreises, der diese Maßnahmen durchführt. Wir plädieren hier für die Nutzung der bestehenden personellen Ressourcen – z.B. Sprachheilpädagoginnen und -pädagogen, Sprachtherapeutinnen und -therapeuten und klinische Linguisten sind ausgebildete Fachkräfte mit den entsprechenden Kompetenzen.

Welche Aspekte müssen unbedingt beachtet werden?

Die derzeit in Deutschland Anwendung findenden Verfahren zur Feststellung des Sprachstandes – z.B. DEUTSCH PLUS (Berlin), HAVAS 4 und 5 (Hamburg), FIT IN DEUTSCH (Niedersachsen), DELFIN 4 (Nordrhein-Westfalen), SSMIK (Bayern) – genügen unseren Ansprüchen an ein aus-sagekräftiges Verfahren zur vorschulischen Sprachstandserhebung nicht. Bei der **Konzeption eines Sprachstandserhebungsverfahrens** müssen diverse Kriterien einbezogen werden, um ein Mindestmaß an Verlässlichkeit und Aussagekraft zu bieten (vgl. Lüdtke, Kallmeyer 2007a/b, Kallmeyer 2007):

Berücksichtigung aller sprachlichen Komponenten (phonetisch-phonologische, morphologisch-syntaktische, lexikalisch-semantische und kommunikativ-pragmatische Aspekte; sprachproduktive und sprachrezeptive Fähigkeiten; meta-sprachliche und meta-kommunikative Kompetenzen; spezifische Sprachwahrnehmungsleistungen)

Einbeziehung aktueller Forschungsergebnisse bei der Konzeption

Explizierung des zugrunde liegenden Sprachbegriffs sowie der *Spracherwerbstheorie förderdiagnostische Ausrichtung* durch die Möglichkeit einer *qualitativen Fehleranalyse*

Evaluation der Güte

Berücksichtigung der spezifischen Spracherwerbssituation von monolingual deutschen bzw. bilingualen Kindern durch eine *differenzierte Auswertung*

Betrachtung beider Sprachen bei bilingual aufwachsenden Kindern und In-Beziehung-Setzen dieser Ergebnisse

Anzustreben ist zudem unbedingt eine Aussage bezüglich der **Differenzierung von Sprachförderung und Sprachtherapie** durch:

Einsatz von Erhebungsinstrumenten, die Kinder mit sprachtherapiebedürftigen Sprachstörungen frühzeitig identifizieren
 Umsetzung entsprechender sprachstörungsspezifischer Therapiemethoden & -konzepte
 Implementierung von integrierten, interdisziplinären Kompetenzteams der Sprachförderung und Sprachtherapie in vorschulischen Bildungsinstitutionen
 Entwicklung von Elternberatungs- und Fortbildungsstrukturen zur Diagnostik und Therapie von Sprachstörungen

An einer **Qualitätsentwicklung und -sicherung der organisatorischen Umsetzung** sollten die Vertreter der Bildungspolitik ein besonderes Interesse haben. *dgs* und *dbs* schlagen daher die *Durchführung wissenschaftlicher Begleitforschung* (wie bei PISA) vor. Ferner sollte von Seiten der Praxis und der Wissenschaft ein *Nachweis über Effektivität und Effizienz der Maßnahmen* erhoben und erbracht werden.

Ein Plädoyer für Chancengleichheit in ganz Deutschland!

Für erstrebenswert halten wir letztlich ein von Fachvertretern und Bildungspolitikern gemeinsam vertretenes, *bundesweit einheitliches Konzept zur vorschulischen Sprachstandserhebung und Sprachförderung* mit eindeutigen gesetzlichen Regelungen. Auch wenn Zentralisierungsbestrebungen immer äußerst kritisch zu reflektieren sind, würde ein diesbezüglicher Abbau des föderalistischen Bildungssystems der Bundesrepublik Deutschland die organisatorische Umsetzung der Maßnahmen um ein erhebliches Maß erleichtern und voraussichtlich zu einer konsequenteren, zielgenaueren und effektiveren Ausführung beitragen. Ein großer Schritt in Richtung Chancengleichheit für Kinder mit und ohne Migrationshintergrund würde so im gesamten deutschen Bildungssystem gemacht.

Literatur

- BMBF* (Bundesministerium für Bildung und Forschung, Kultusministerkonferenz der Länder) (Hrsg.) (2006): Aktivitäten im Bereich Benachteiligtenförderung und Förderung von Migranten. URL: http://www.bmbf.de/pub/migration_aktivitaeten.pdf, 22.10.2007.
- De Langen, U. / Mahack, V. (2007): Sprachtherapie ist kein Kinderspiel (Informationsbroschüre). Herausgeber: Deutscher Bundesverband der akademischen Sprachtherapeuten e.V. (dbs) / Selbstverlag.
- Kallmeyer, K. (2007): Vorschulische Maßnahmen zur Sprachstandserhebung in den deutschen Bundesländern - Eine wissenschaftliche Synopse ausgewählter praxisrelevanter Verfahren. Aachen: Shaker.
- Lüdtke, U., Kallmeyer, K. (2007a): Vorschulische Maßnahmen zur Sprachstandserhebung und Sprachförderung in den deutschen Bundesländern: Wissenschaftliche Vorschläge zur Optimierung bildungspolitischer Initiativen. In: Die Sprachheilarbeit 52 (6), 244-260.
- Lüdtke, U., Kallmeyer, K. (2007b): Kritische Analyse ausgewählter Sprachstandserhebungsverfahren für Kinder vor Schuleintritt aus Sicht der Linguistik, Diagnostik und Mehrsprachigkeitsforschung. In: Die Sprachheilarbeit 52 (6), 261-278.

Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit oder Teile davon wurden bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde als Dissertation vorgelegt. Ferner erkläre ich, dass ich nicht bereits eine gleichartige Doktorprüfung an einer Hochschule endgültig nicht bestanden habe.

Erfurt, 01.11.2011

Sandra Kuhn

Lebenslauf

Oktober 2001 bis Februar 2007 – Diplom Psychologie, Universität des Saarlandes

Oktober 2003 bis August 2007 – B.Sc.Computerlinguistik, Universität des Saarlandes

Oktober 2008 bis Oktober 2011 – Promotion Psychologie, Universität Erfurt

Studien – und Forschungsaufenthalte: Trinity College Dublin, Irland (2003-2004); University of New England, Armidale, Australien (2004)